File No. BC45300-1US
Serial No.: 10/650,608

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE ACTION

In re Application of: Jean-Paul Cassart, et al.
Serial No.: 10/650,608
Filing Date: 28 Aug 2003
For: TUMOUR SPECIFIC ANIMAL PROTEINS

Examiner: Min-Tam Davis
Art Unit: 1642
Confirmation No: 8978

Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

COMMUNICATION

Ma'am:

This Communication is being filed contemporaneously with courtesy copies of the certified copies of the priority applications of record in PCT/EP01/01779, WO01/62778, the benefit of which is claimed in the present case.

Remarks begin on page **2** of this paper.

REMARKS

Applicants submit herewith a copy of the following:

1) A copy of the certified copy of Great Britain Patent Application No. 0004269.7 filed February 23, 2000;

2) A copy of the certified copy of Great Britain Patent Application No. 0009905.1 filed April 20, 2000; and

3) A copy of the certified copy of Great Britain Patent Application No. 0021080.7 filed August 25, 2000;

each of which are of record in PCT/EP01/01779.

The Commissioner is hereby authorized to charge any fees required or credit any overpayment to Deposit Account No. 07-1392.

Respectfully submitted,

/Eric J. Kron/
Eric J. Kron
Attorney for Applicant
Registration No. 45,941

Date: 20 May 09
GlaxoSmithKline
Corporate Intellectual Property
Five Moore Drive
P.O. Box 13398
Research Triangle Park, NC 27709-3398
Phone: 919-483-8961
Facsimile: 919-483-7988

EPA/EPO/OEB
D-80298 München
☎ +49 89 2399-0
TX 523 656 epmu d
FAX +49 89 2399-4465

Europäisches Patentamt
Eingangsstelle

European Patent Office
Receiving Section

Office européen des brevets
Section de Dépôt

WIPO
The International Bureau
34, Chemin des Colombettes
CH-1211 GENEVA 20
SWITZERLAND

EPO - DG 1
15. 11. 2002
(104)

Datum/Date
05-11-2002

Zeichen/Ref./Réf.	Anmeldung Nr./Application No./Demande n°./Patent Nr./Patent No./Brevet n°.
KCR/BC45300	PCT/EP0101779 - EP/01929345.5-2401 /

Anmelder/Applicant/Demandeur/Patentinhaber/Proprietor/Titulaire
SMITHKLINE BEECHAM BIOLOGICALS S.A.

For the aforementioned application, you are hereby notified that the EPO has not received:

( ) a) the notification of election (Rule 61.2 PCT)

( ) b) the copy of the international preliminary examination report (Art. 36(3)(a) PCT)

(X) c) the ~~XXX~~ (copies) of the priority document(s) (Rule 17.2(a) PCT).

( ) ..... .

You are kindly requested to forward the missing item at your earliest convenience to the EPO in its capacity as designated/elected Office.

RECEIVING SECTION
Adams, Timothy

Answer of the International Bureau [IB]:

( ) The requested item [a), b) or c)] is not available with the IB.

Reason: ..... .

( ) The applicant has requested the receiving Office to issue a priority document [c)] pursuant to Rule 17.1(b) PCT, but the IB did not receive it.

The International Bureau

PCT/EP 0 1 / 0 1 7 7 9

EPO - DG 1

0 9. 07. 2001

(83)

REC'D 2 3 AUG 2001

WIPO PCT

The Patent Office
Concept House
Cardiff Road
Newport
South Wales
NP10 8QQ

PRIORITY DOCUMENT

SUBMITTED OR TRANSMITTED IN COMPLIANCE WITH RULE 17.1(a) OR (b)

I, the undersigned, being an officer duly authorised in accordance with Section 74(1) and (4) of the Deregulation & Contracting Out Act 1994, to sign and issue certificates on behalf of the Comptroller-General, hereby certify that annexed hereto is a true copy of the documents as originally filed in connection with the patent application identified therein.

In accordance with the Patents (Companies Re-registration) Rules 1982, if a company named in this certificate and any accompanying documents has re-registered under the Companies Act 1980 with the same name as that with which it was registered immediately before re-registration save for the substitution as, or inclusion as, the last part of the name of the words "public limited company" or their equivalents in Welsh, references to the name of the company in this certificate and any accompanying documents shall be treated as references to the name with which it is so re-registered.

In accordance with the rules, the words "public limited company" may be replaced by p.l.c., plc, P.L.C. or PLC.

Re-registration under the Companies Act does not constitute a new legal entity but merely subjects the company to certain additional company law rules.

Signed

Dated 31 May 2001

Patents Form 1/77

Patents Act 1977
(Rule 16)

The Patent Office

1/77

25APR00 E531764-2 002029
P01/7700 0.00-0009905.1

Request for grant of a patent

(See the notes on the back of this form. You can also get an explanatory leaflet from the Patent Office to help you fill in this form)

THE PATENT OFFICE
20 APR 2000
LONDON

The Patent Office
Cardiff Road
Newport
Gwent NP9 1RH

1. Your reference — AWRT/KLP/RSH/BC45306

2. Patent ap *(The Patent* — 0009905.1 — 20 APR 2000

3. Full name, address and postcode of the or of each applicant *(underline all surnames)*

   Patents ADP number *(if you know it)*

   If the applicant is a corporate body, give the country/state of its incorporation

   SmithKline Beecham Biologicals s.a.
   rue de l'Institut 89, B-1330 Rixensart, Belgium

   578111760

   Belgium

4. Title of the invention — NOVEL COMPOUNDS

5. Name of your agent *(if you have one)*

   "Address for service" in the United Kingdom to which all correspondence should be sent *(including the postcode)*

   Patents ADP number *(if you know it)*

   CORPORATE INTELLECTUAL PROPERTY

   SMITHKLINE BEECHAM PLC
   TWO NEW HORIZONS COURT
   BRENTFORD
   MIDDLESEX TW8 9EP

   58009 74004

6. If you are declaring priority from one or more earlier patent applications, give the country and the date of filing of the or each of these earlier applications and *(if you know it)* the or each application number

Country	Priority application number *(if you know it)*	Date of filing *(day / month / year)*

7. If this application is divided or otherwise derived from an earlier UK application, give the number and the filing date of the earlier application

Number of earlier application	Date of filing *(day / month / year)*

8. Is a statement of inventorship and of right to grant of a patent required in support of this request? *(Answer yes if:*
   a) any applicant named in part 3 is not an inventor, or
   b) there is an inventor who is named as an applicant, or
   c) any named applicant is a corporate body
   See note (d)

Patents Form 1/77

9. Enter the number of sheets for any of the following items you are filing with this form. Do not count copies of the same document

Continuation sheets of this form	
Description	25
Claim(s)	1
Abstract	1
Drawings	2

10. If you are also filing any of the following, state how many against each item.

Priority Documents	
Translations of priority documents	
Statement of inventorship and right to grant of a patent *(Patents Form 1/77)*	
Request for preliminary examination and search *(Patents Form 9/77)*	
Request for substantive examination *(Patents Form 10/77)*	
Any other documents *(please specify)*	

11. We request the grant of a patent on the basis of this application
Signature A W R TYRRELL Date 20-Apr-00

12. Name and daytime telephone number of person to contact in the United Kingdom: A W R TYRRELL 020 89752585

Warning

After an application for a Patent has beeen filed, the Comptroller of the Patent Office will consider wheiher publication or communication of the invention should be prohibited or restricted under Section 22 of the Patents Act 1977. You will be informed if it is necessary to prohibit or restrict your invention in this way. Furthermore, if you live in the United Kingdom, Section 23 of the Patents Act 1977 stops you from applying for a patent abroad without first getting written permission unless an application has been filed at least six weeks beforehand in the United Kingdom for a patent for the same invention and either no direction prohibiting publication or communication has been given, or any such direction has been revoked.

Notes

a) If you need help to fill in this form or you have any questions, please contact the Patent Office on 0645 500505
b) Write your answers in capital letters using black ink or you may type them.
c) If there is not enough space for all relevant details on any part of this form, please continue on a separate sheet of paper and write "see continuation sheet" in the relevant part(s). Any continuation sheet should be attached to this form.
d) If you have answered 'Yes' Patents Form 7/77 will need to be filed.

Novel Compounds

Field of the Invention

This invention relates to newly identified polypeptides and polynucleotides encoding such polypeptides, to their use in diagnostics, prophylaxis and therapy and in identifying compounds which may be agonists, antagonists and /or inhibitors which are potentially useful in therapy, and to production of such polypeptides and polynucleotides.

Background of the Invention

The drug discovery process is currently undergoing a fundamental revolution as it embraces 'functional genomics', that is, high throughput genome- or gene-based biology. This approach is rapidly superceding earlier approaches based on 'positional cloning'. A phenotype, that is a biological function or genetic disease, would be identified and this would then be tracked back to the responsible gene, based on its genetic map position.

Functional genomics relies heavily on high-throughput DNA sequencing technologies and the various tools of bioinformatics to identify gene sequences of potential interest from the many molecular biology databases now available. cDNA libraries enriched for genes of relevance to a particular tissue or physiological situation can be constructed using recently developed subtractive cloning strategies. Construction of subtractive libraries reduces cDNA sequence complexity, focuses experimental resources on relevant genes and thus greatly accelerates the gene identification process. There is a continuing need to identify and characterise further genes and their related polypeptides/proteins, as targets for drug discovery.

High throughput genome- or gene-based biology allows new approaches to the identification and cloning of target genes for useful immune responses for the prevention and vaccine therapy of diseases such as cancer and autoimmunity.

Summary of the Invention

The present invention relates to CASB7439, in particular CASB7439 polypeptides and CASB7439 polynucleotides, recombinant materials and methods for their production. In another aspect, the invention relates to methods for using such polypeptides and polynucleotides, including the treatment of cancer and autoimmune diseases, hereinafter referred to as "the Diseases", amongst others. In a further aspect, the invention relates to methods for identifying agonists and antagonists/inhibitors using the materials provided by the invention, and treating conditions associated with CASB7439 imbalance with the identified compounds In a still further aspect, the invention relates to diagnostic assays for detecting diseases associated with inappropriate CASB7439 activity or levels.

Description of the Invention

In a first aspect, the present invention relates to CASB7439 polypeptides. Such peptides include isolated polypetides comprising an amino acid sequence which has at least 70%identity, preferably at least 80% identity, more preferably at least 90% identity, yet more preferably at least 95% identity, most preferably at least 97-99% identity, to that of SEQ ID NO:2 AND SEQ ID NO:3 over the entire length of SEQ ID NO:2 AND SEQ ID NO:3 . Such polypeptides include those comprising the amino acid of SEQ ID NO:2 AND SEQ ID NO:3 .

Further peptides of the present invention include isolated polypeptides in which the amino acid sequence has at least 70% identity, preferably at least 80% identity, more preferably at least 90% identity, yet more preferably at least 95% identity, most preferably at least 97-99% identity, to the amino acid sequence of SEQ ID NO:2 AND SEQ ID NO:3 over the entire length of SEQ ID NO:2 AND SEQ ID NO:3 . Such polypeptides include the polypeptide of SEQ ID NO:2 AND SEQ ID NO:3 .

Further peptides of the present invention include isolated polypeptides encoded by a polynucleotide comprising the sequence contained in SEQ ID NO:1.

Polypeptides and polynucleotides of the present invention are believed to be important immunogens for specific prophylactic or therapeutic immunization against tumors, because they are highly overexpressed in tumors and can thus be targeted by antigen-specific immune mechanisms leading to the destruction of the tumor cell. They can also be used to diagnose the occurrence of tumor cells. Furthermore, their inappropriate expression in certain circumstances can cause an induction of autoimmune, inappropriate immune responses, which could be corrected through appropriate vaccination using the same polypeptides or polynucleotides. In this respect the most important biological activities to our purpose are the antigenic and immunogenic activities of the polypeptide of the present invention. A polypeptide of the present invention may also exhibit at least one other biological activity of CASB7439, which could qualify it as a target for therapeutic or prophylactic intervention different from that linked to the immune response.

The polypeptides of the present invention may be in the form of the "mature" protein or may be a part of a larger protein such as a fusion protein. It is often advantageous to include an additional amino acid sequence which contains secretory or leader sequences, pro-sequences, sequences which aid in purification such as multiple histidine residues, or an additional sequence for stability during recombinant production. Furthermore, addition of exogenous polypeptide or lipid tail or polynucleotide sequences to increase the immunogenic potential of the final molecule is also considered.

The present invention also includes variants of the aforementioned polypeptides, that is polypeptides that vary from the referents by conservative amino acid substitutions, whereby a residue

is substituted by another with like characteristics. Typical such substitutions are among Ala, Val, Leu and Ile; among Ser and Thr; among the acidic residues Asp and Glu; among Asn and Gln; and among the basic residues Lys and Arg; or aromatic residues Phe and Tyr. Particularly preferred are variants in which several, 5-10, 1-5, 1-3, 1-2 or 1 amino acids are substituted, deleted, or added in any combination.

Polypeptides of the present invention can be prepared in any suitable manner. Such polypeptides include isolated naturally occurring polypeptides, recombinantly produced polypeptides, synthetically produced polypeptides, or polypeptides produced by a combination of these methods. Means for preparing such polypeptides are well understood in the art.

In a further aspect, the present invention relates to CASB7439 polynucleotides. Such polynucleotides include isolated polynucleotides comprising a nucleotide sequence encoding a polypeptide which has at least 70% identity, preferably at least 80% identity, more preferably at least 90% identity, yet more preferably at least 95% identity, to the amino acid sequence ofSEQ ID NO:2 AND SEQ ID NO:3 , over the entire length of SEQ ID NO:2 AND SEQ ID NO:3 . In this regard, polypeptides which have at least 97% identity are highly preferred, whilst those with at least 98-99% identity are more highly preferred, and those with at least 99% identity are most highly preferred. Such polynucleotides include a polynucleotide comprising the nucleotide sequence contained in SEQ ID NO:1 encoding the polypeptide of SEQ ID NO:2 AND SEQ ID NO:3 .

Further polynucleotides of the present invention include isolated polynucleotides comprising a nucleotide sequence that has at least 70% identity, preferably at least 80% identity, more preferably at least 90% identity, yet more preferably at least 95% identity,to a nucleotide sequence encoding a polypeptide of SEQ ID NO:2 AND SEQ ID NO:3 , over the entire coding region. In this regard, polynucleotides which have at least 97% identity are highly preferred, whilst those with at least 98-99% identity are more highly preferred, and those with at least 99% identity are most highly preferred.

Further polynucleotides of the present invention include isolated polynucleotidescomprising a nucleotide sequence which has at least 70% identity, preferably at least 80% identity, more preferably at least 90% identity, yet more preferably at least 95% identity, to SEQ ID NO:1 over the entire length of SEQ ID NO:1. In this regard, polynucleotides which have at least 97% identity are highly preferred, whilst those with at least 98-99% idcntiy are more highly preferred, and those with at least 99% identity are most highly preferred. Such polynucleotides include a polynucleotide comprising the polynucleotide of SEQ ID NO:1 as well as the polynucleotide of SEQ ID NO:1. Said polynucleotide can be inserted in a suitable plasmid or recombinant microrganism vector and used for immunization (see for example Wolff et. al., Science 247:1465-1468 (1990); Corr et. al., J. Exp. Med. 184:1555-1560 (1996); Doe et. al., Proc. Natl. Acad. Sci. 93:8578-8583 (1996)).

The invention also provides polynucleotides which are complementary to all the above described polynucleotides.

The nucleotide sequence of SEQ ID NO:1 shows homology with Homo sapiens achaete-scute complex (Drosophila) homolog-like 2 (ASCL2) mRNA (genbank accession NM_005170). The nucleotide sequence of SEQ ID NO:1 is a cDNA sequence and comprises a polypeptide encoding sequence (nucleotide 547 to 1125) encoding a polypeptide of 193 amino acids, the polypeptide of SEQ ID NO:2. The nucleotide sequence encoding the polypeptide of SEQ ID NO:2 may be identical to the polypeptide encoding sequence contained in SEQ ID NO:1 or it may be a sequence other than the one contained in SEQ ID NO:1, which, as a result of the redundancy (degeneracy) of the genetic code, also encodes the polypeptide of SEQ ID NO:2. The polypeptide of the SEQ ID NO:2 is structurally related to other proteins of the Achaete-Scute protein family, having homology and/or structural similarity with human Achaete-Scute homologue 2 (genbank accession AAB86993).

Human Achaete Scute Homologue 2 (HASH2) gene, officially designated ASCL2 (Achaete Scute complex like 2) is a homologue of the Drosophila Achaete and Scute genes. HASH2 is expressed in the extravillus trophoblasts of the developing placenta only, and maps on chromosome 11p15 close to IGF2 and H19.

The mouse achaete-scute homolog-2 gene (MASH2) encodes a transcription factor playing a role in the development of the trophoblast.

The Ascl2 gene is paternally imprinted in the mouse, and the lack of HASH2 expression in non-malignant hydatidiform (androgenetic) moles indicates that HASH2 is also imprinted in man.

The nucleotide sequence of SEQ ID NO:1 also comprises an another polypeptide encoding sequence (nucleotide 1483 to 401) encoding a polypeptide of 361 amino acids, the polypeptide of SEQ ID NO:3. The nucleotide sequence encoding the polypeptide of SEQ ID NO:3 may be identical to the polypeptide encoding sequence contained in SEQ ID NO:1 or it may be a sequence other than the one contained in SEQ ID NO:1, which, as a result of the redundancy (degeneracy) of the genetic code, also encodes the polypeptide of SEQ ID NO:3. The polypeptide of the SEQ ID NO:2 is structurally related to other proteins of the splicing coactivator protein family, having homology and/or structural similarity with AAF21439 Aaf21439 homo sapiens splicing coactivator subunit srm300.

Preferred polypeptides and polynucleotides of the present invention are expected to have, *inter alia*, similar biological functions/properties to their homologous polypeptides and polynucleotides. Furthermore, preferred polypeptides and polynucleotides of the present invention have at least one CASB7439 activity.

The present invention also relates to partial or other polynucleotide and polypeptide sequences which were first identified prior to the determination of the corresponding full length sequences of SEQ ID NO:1, SEQ ID NO:2 and SEQ ID NO:3.

Accordingly, in a further aspect, the present invention provides for an isolated polynucleotide comprising:

(a) a nucleotide sequence which has at least 70% identity, preferably at least 80% identity, more preferably at least 90% identity, yet more preferably at least 95% identity, even more preferably at least 97-99% identity to SEQ ID NO:4 over the entire length of SEQ ID NO:4;

(b) a nucleotide sequence which has at least 70% identity, preferably at least 80% identity, more preferably at least 90% identity, yet more preferably at least 95% identity, even more preferably at least 97-99% identity, to SEQ ID NO:1 over the entire length of SEQ ID NO:4;

(c) the polynucleotide of SEQ ID NO:4;

as well as the polynucleotide of SEQ ID NO:4.

The nucleotide sequence of SEQ ID NO:4 and the peptide sequence encoded thereby are derived from EST (Expressed Sequence Tag) sequences. It is recognised by those skilled in the art that there will inevitably be some nucleotide sequence reading errors in EST sequences (see Adams, M.D. *et al*, Nature 377 (supp) 3, 1995). Accordingly, the nucleotide sequence of SEQ ID NO:4 and the peptide sequence encoded therefrom are therefore subjec to the same inherent limitations in sequence accuracy. Furthermore, the peptide sequence encoded by SEQ ID NO:5 comprises a region of identity or close homology and/or close structural similarity (for example a conservative amino acid difference) with the closest homologous or structurally similar protein.

Polynucleotides of the present invention may be obtained, using standard cloning and screening techniques, from a cDNA library derived from mRNA in cells ofhuman colorectal and germ cell tumors, using the expressed sequence tag (EST) analysis (Adams, M.D., *et al.* Science (1991) 252:1651-1656; Adams, M.D. *et al.*, Nature, (1992) *355*:632-634; Adams, M.D., *et al.*, Nature (1995) 377 Supp:3-174). Polynucleotides of the invention can also be obtained from natural sources such as genomic DNA libraries or can be synthesized using well known and commercially available techniques.

When polynucleotides of the present invention are used for the recombinant production of polypeptides of the present invention, the polynucleotide may include the coding sequence for the mature polypeptide, by itself; or the coding sequence for the mature polypeptide in reading frame with other coding sequences, such as those encoding a leader or secretory sequence, a pre-, or pro- or prepro- protein sequence, or other fusion peptide portions. For example, a marker sequence which facilitates purification of the fused polypeptide can be encoded. In certain preferred embodiments of this aspect of the invention, the marker sequence is a hexa-histidine peptide, as provided in the pQE

vector (Qiagen, Inc.) and described in Gentz *et al.*, Proc Natl Acad Sci USA (1989) 86:821-824, or is an HA tag. The polynucleotide may also contain non-coding 5' and 3' sequences, such as transcribed, non-translated sequences, splicing and polyadenylation signals, ribosome binding sites and sequences that stabilize mRNA.

Further embodiments of the present invention include polynucleotides encoding polypeptide variants which comprise the amino acid sequence of SEQ ID NO:2 and in which several, for instance from 5 to 10, 1 to 5, 1 to 3, 1 to 2 or 1, amino acid residues are substituted, deleted or added, in any combination.

Polynucleotides which are identical or sufficiently identical to a nucleotide sequence contained in SEQ ID NO:1, may be used as hybridization probes for cDNA and genomic DNA or as primers for a nucleic acid amplification (PCR) reaction, to isolate full-length cDNAs and genomic clones encoding polypeptides of the present invention and to isolate cDNA and genomic clones of other genes (including genes encoding homologs and orthologs from species other than human) that have a high sequence similarity to SEQ ID NO:1. Typically these nucleotide sequences are 70% identical, preferably 80% identical, more preferably 90% identical, most preferably 95% identical to that of the referent. The probes or primers will generally comprise at least 15 nucleotides, preferably, at least 30 nucleotides and may have at least 50 nucleotides. Particularly preferred probes will have between 30 and 50 nucleotides. In particular, polypeptides or polynucleotides derived from sequences from homologous animal origin could be used as immunogens to obtain a cross-reactive immune response to the human gene.

A polynucleotide encoding a polypeptide of the present invention, including homologs and orthologs from species other than human, may be obtained by a process which comprises the steps of screening an appropriate library under stringent hybridization conditions with a labeled probe having the sequence of SEQ ID NO: 1 or a fragment thereof; and isolating full-length cDNA and genomic clones containing said polynucleotide sequence. Such hybridization techniques are well known to the skilled artisan. Preferred stringent hybridization conditions include overnight incubation at 42°C in a solution comprising: 50% formamide, 5xSSC (150mM NaCl, 15mM trisodium citrate), 50 mM sodium phosphate (pH7.6), 5x Denhardt's solution, 10 % dextran sulfate, and 20 microgram/ml denatured, sheared salmon sperm DNA; followed by washing the filters in 0.1x SSC at about 65°C. Thus the present invention also includes polynucleotides obtainable by screening an appropriate library under stingent hybridization conditions with a labeled probe having the sequence of SEQ ID NO:1 or a fragment thereof.

The skilled artisan will appreciate that, in many cases, an isolated cDNA sequence will be incomplete, in that the region coding for the polypeptide is cut short at the 5' end of the cDNA. This is a consequence of reverse transcriptase, an enzyme with inherently low 'processivity' (a measure of the ability of the enzyme to remain attached to the template during the polymerisation

reaction), failing to complete a DNA copy of the mRNA template during 1st strand cDNA synthesis.

There are several methods available and well known to those skilled in the art to obtain full-length cDNAs, or extend short cDNAs, for example those based on the method of Rapid Amplification of cDNA ends (RACE) (see, for example, Frohman et al., PNAS USA 85, 8998-9002, 1988). Recent modifications of the technique, exemplified by the Marathon™ technology (Clontech Laboratories Inc.) for example, have significantly simplified the search for longer cDNAs. In the Marathon™ technology, cDNAs have been prepared from mRNA extracted from a chosen tissue and an 'adaptor' sequence ligated onto each end. Nucleic acid amplification (PCR) is then carried out to amplify the 'missing' 5' end of the cDNA using a combination of gene specific and adaptor specific oligonucleotide primers. The PCR reaction is then repeated using 'nested' primers, that is, primers designed to anneal within the amplified product (typically an adaptor specific primer that anneals further 3' in the adaptor sequence and a gene specific primer that anneals further 5' in the known gene sequence). The products of this reaction can then be analysed by DNA sequencing and a full-length cDNA constructed either by joining the product directly to the existing cDNA to give a complete sequence, or carrying out a separate full-length PCR using the new sequence information for the design of the 5' primer.

Recombinant polypeptides of the present invention may be prepared by processes well known in the art from genetically engineered host cells comprising expression systems. Accordingly, in a further aspect, the present invention relates to expression systems which comprise a polynucleotide or polynucleotides of the present invention, to host cells which are genetically engineered with such expression sytems and to the production of polypeptides of the invention by recombinant techniques. Cell-free translation systems can also be employed to produce such proteins using RNAs derived from the DNA constructs of the present invention.

For recombinant production, host cells can be genetically engineered to incorporate expression systems or portions thereof for polynucleotides of the present invention. Introduction of polynucleotides into host cells can be effected by methods described in many standard laboratory manuals, such as Davis et al, Basic Methods in Molecular Biology (1986) and Sambrook *et al.*, Molecular Cloning: A Laboratory Manual, 2nd Ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. (1989). Preferred such methods include, for instance, calcium phosphate transfection, DEAE-dextran mediated transfection, transvection, microinjection, cationic lipid-mediated transfection, electroporation, transduction, scrape loading, ballistic introduction or infection.

Representative examples of appropriate hosts include bacterial cells, such as *streptococci, staphylococci, E. coli, Streptomyces* and *Bacillus subtilis* cells; fungal cells, such as yeast cells and

Aspergillus cells; insect cells such as *Drosophila* S2 and *Spodoptera* Sf9 cells; animal cells such as CHO, COS, HeLa, C127, 3T3, BHK, HEK 293 and Bowes melanoma cells; and plant cells.

A great variety of expression systems can be used, for instance, chromosomal, episomal and virus-derived systems, e.g., vectors derived from bacterial plasmids, from bacteriophage, from transposons, from yeast episomes, from insertion elements, from yeast chromosomal elements, from viruses such as baculoviruses, papova viruses, such as SV40, vaccinia viruses, adenoviruses, fowl pox viruses, pseudorabies viruses and retroviruses, and vectors derived from combinations thereof, such as those derived from plasmid and bacteriophage genetic elements, such as cosmids and phagemids. The expression systems may contain control regions that regulate as well as engender expression. Generally, any system or vector which is able to maintain, propagate or express a polynucleotide to produce a polypeptide in a host may be used. The appropriate nucleotide sequence may be inserted into an expression system by any of a variety of well-known and routine techniques, such as, for example, those set forth in Sambrook *et al.*, Molecular Cloning, A Laboratory Manual (supra). Appropriate secretion signals may be incorporated into the desired polypeptide to allow secretion of the translated protein into the lumen of the endoplasmic reticulum, the periplasmic space or the extracellular environment. These signals may be endogenous to the polypeptide or they may be heterologous signals.

If a polypeptide of the present invention is to be expressed for use in screening assays, it is generally preferred that the polypeptide be produced at the surface of the cell. In this event, the cells may be harvested prior to use in the screening assay. If the polypeptide is secreted into the medium, the medium can be recovered in order to recover and purify the polypeptide. If produced intracellularly, the cells must first be lysed before the polypeptide is recovered.

Polypeptides of the present invention can be recovered and purified from recombinant cell cultures by well-known methods including ammonium sulfate or ethanol precipitation, acid extraction, anion or cation exchange chromatography, phosphocellulose chromatography, hydrophobic interaction chromatography, affinity chromatography, hydroxylapatite chromatography and lectin chromatography. Most preferably, high performance liquid chromatography is employed for purification. Well known techniques for refolding proteins may be employed to regenerate active conformation when the polypeptide is denatured during isolation and or purification.

Another important aspect of the invention relates to a method for inducing , re-inforcing or modulating an immunological response in a mammal which comprises inoculating the mammal with a fragment or the entire polypeptide or polynucleotide of the present invention, adequate to produce antibody and/or T cell immune response for prophylaxis or for therapeutic treatment of the diseases hereinbefore mentioned, amongst others. Yet another aspect of the invention relates to a method of inducing, re-inforcing or modulating immunological response in a mammal which comprises, delivering a polypeptide of the present invention *via* a vector or cell directing

expression of the polynucleotide and coding for the polypeptide *in vivo* in order to induce such an immunological response to produce immune responses for prophylaxis or treatment of said mammal from diseases.

A further aspect of the invention relates to an immunological/vaccine formulation (composition) which, when introduced into a mammalian host, induces, re-inforces or modulates an immunological response in that mammal to a polypeptide of the present invention wherein the composition comprises a fragment or the entire polypeptide or polynucleotide of the present invention. The vaccine formulation may further comprise a suitable carrier. Since a polypeptide may be broken down in the stomach, it is preferably administered parenterally (for instance, subcutaneous, intramuscular, intravenous, or intradermal injection). Formulations suitable for parenteral administration include aqueous and non-aqueous sterile injection solutions which may contain anti-oxidants, buffers, bacteriostats and solutes which render the formulation isotonic with the blood of the recipient; and aqueous and non-aqueous sterile suspensions which may include suspending agents or thickening agents. The formulations may be presented in unit-dose or multi-dose containers, for example, sealed ampoules and vials and may be stored in a freeze-dried condition requiring only the addition of the sterile liquid carrier immediately prior to use. The vaccine formulation may also include adjuvant systems for enhancing the immunogenicity of the formulation, such as oil-in water systems, immunomodulators and other systems known in the art. The dosage will depend on the specific activity of the vaccine and can be readily determined by routine experimentation.

A further aspect of the invention relates to the *in vitro* induction of immune responses to a fragment or the entire polypeptide or polynucleotide of the present invention or a molecule comprising the polypeptide or polynucleotide of the present invention, using cells from the immune system of a mammal, and reinfusing these activated immune cells of the mammal for the treatment of disease. Activation of the cells from the immune system is achieved by in vitro incubation with the entire polypeptide or polynucleotide of the present invention or a molecule comprising the polypeptide or polynucleotide of the present invention in the presence or absence of various immunomodulator molecules.

A further aspect of the invention relates to the immunization of a mammal by administration of antigen presenting cells modified by *in vitro* loading with part or the entire polypeptide of the present invention or a molecule comprising the polypeptide of the present invention and administered *in vivo* in an immunogenic way. Alternatively, antigen presenting cells can be transfected *in vitro* with a vector containing a fragment or the entire polynucleotide of the present invention or a molecule comprising the polynucleotide of the present invention, such as to express the corresponding polypeptide, and administered *in vivo* in an immunogenic way.

This invention also relates to the use of polynucleotides of the present invention as diagnostic reagents. Detection of a mutated form of the gene characterised by the polynucleotide of SEQ ID NO:1 which is associated with a dysfunction will provide a diagnostic tool that can add to, or define, a diagnosis of a disease, or susceptibility to a disease, which results from under-expression, over-expression or altered expression of the gene. Individuals carrying mutations in the gene may be detected at the DNA level by a variety of techniques.

Nucleic acids for diagnosis may be obtained from a subject's cells, such as from blood, urine, saliva, tissue biopsy or autopsy material. The genomic DNA may be used directly for detection or may be amplified enzymatically by using PCR or other amplification techniques prior to analysis. RNA or cDNA may also be used in similar fashion. Deletions and insertions can be detected by a change in size of the amplified product in comparison to the normal genotype. Point mutations can be identified by hybridizing amplified DNA to labeled CASB7439 nucleotide sequences. Perfectly matched sequences can be distinguished from mismatched duplexes by RNase digestion or by differences in melting temperatures. DNA sequence differences may also be detected by alterations in electrophoretic mobility of DNA fragments in gels, with or without denaturing agents, or by direct DNA sequencing (ee, e.g., Myers *et al.*, Science (1985) 230:1242). Sequence changes at specific locations may also be revealed by nuclease protection assays, such as RNase and S1 protection or the chemical cleavage method (see Cotton *et al.*, Proc Natl Acad Sci USA (1985) 85: 4397-4401). In another embodiment, an array of oligonucleotides probes comprising CASB7439 nucleotide sequence or fragments thereof can be constructed to conduct efficient screening of e.g., genetic mutations. Array technology methods are well known and have general applicability and can be used to address a variety of questions in molecular genetics including gene expression, genetic linkage, and genetic variability (see for example: M.Chee et al., Science, Vol 274, pp 610-613 (1996)).

The diagnostic assays offer a process for diagnosing or determining a susceptibility to the Diseases through detection of mutation in the CASB7439 gene by the methods described. In addition, such diseases may be diagnosed by methods comprising determining from a sample derived from a subject an abnormally decreased or increased level of polypeptide or mRNA. Decreased or increased expression can be measured at the RNA level using any of the methods well known in the art for the quantitation of polynucleotides, such as, for example, nucleic acid amplification, for instance PCR, RT-PCR, RNase protection, Northern blotting and other hybridization methods. Assay techniques that can be used to determine levels of a protein, such as a polypeptide of the present invention, in a sample derived from a host are well-known to those of skill in the art. Such assay methods include radioimmunoassays, competitive-binding assays, Western Blot analysis and ELISA assays.

Thus in another aspect, the present invention relates to a diagnostic kit which comprises:

(a) a polynucleotide of the present invention, preferably the nucleotide sequence of SEQ ID NO: 1, or a fragment thereof;

(b) a nucleotide sequence complementary to that of (a);

(c) a polypeptide of the present invention, preferably the polypeptide of SEQ ID NO:2 or SEQ ID NO:3 or a fragment thereof; or

(d) an antibody to a polypeptide of the present invention, preferably to the polypeptide of SEQ ID NO:2 2 or SEQ ID NO:3.

It will be appreciated that in any such kit, (a), (b), (c) or (d) may comprise a substantial component. Such a kit will be of use in diagnosing a disease or suspectability to a disease, particularly, amongst others.

The nucleotide sequences of the present invention are also valuable for chromosome identification. The sequence is specifically targeted to, and can hybridize with, a particular location on an individual human chromosome. The mapping of relevant sequences to chromosomes according to the present invention is an important first step in correlating those sequences with gene associated disease. Once a sequence has been mapped to a precise chromosomal location, the physical position of the sequence on the chromosome can be correlated with genetic map data. Such data are found in, for example, V. McKusick, Mendelian Inheritance in Man (available on-line through Johns Hopkins University Welch Medical Library). The relationship between genes and diseases that have been mapped to the same chromosomal region are then identified through linkage analysis (coinheritance of physically adjacent genes).

The differences in the cDNA or genomic sequence between affected and unaffected individuals can also be determined. If a mutation is observed in some or all of the affected individuals but not in any normal individuals, then the mutation is likely to be the causative agent of the disease.

The polypeptides of the invention or their fragments or analogs thereof, or cells expressing them, can also be used as immunogens to produce antibodies immunospecific for polypeptides of the present invention. The term "immunospecific" means that the antibodies have substantially greater affinity for the polypeptides of the invention than their affinity for other related polypeptides in the prior art.

Antibodies generated against polypeptides of the present invention may be obtained by administering the polypeptides or epitope-bearing fragments, analogs or cells to an animal, preferably a non-human animal, using routine protocols. For preparation of monoclonal antibodies, any technique which provides antibodies produced by continuous cell line cultures can be used. Examples include the hybridoma technique (Kohler, G. and Milstein, C., Nature (1975) 256:495-497), the trioma technique, the human B-cell hybridoma technique (Kozbor *et al.*, Immunology

Today (1983) 4:72) and the EBV-hybridoma technique (Cole *et al.*, Monoclonal Antibodies and Cancer Therapy, 77-96, Alan R. Liss, Inc., 1985).

Techniques for the production of single chain antibodies, such as those described in U.S. Patent No. 4,946,778, can also be adapted to produce single chain antibodies to polypeptides of this invention. Also, transgenic mice, or other organisms, including other mammals, may be used to express humanized antibodies.

The above-described antibodies may be employed to isolate or to identify clones expressing the polypeptide or to purify the polypeptides by affinity chromatography.

Antibodies against polypeptides of the present invention may also be employed to treat the Diseases, amongst others.

In a further aspect, the present invention relates to genetically engineered soluble fusion proteins comprising a polypeptide of the present invention, or a fragment thereof, and various portions of the constant regions of heavy or light chains of immunoglobulins of various subclasses (IgG, IgM, IgA, IgE). Preferred as an immunoglobulin is the constant part of the heavy chain of human IgG, particularly IgG1, where fusion takes place at the hinge region. In a particular embodiment, the Fc part can be removed simply by incorporation of a cleavage sequence which can be cleaved with blood clotting factor Xa. Furthermore, this invention relates to processes for the preparation of these fusion proteins by genetic engineering, and to the use thereof for drug screening, diagnosis and therapy. A further aspect of the invention also relates to polynucleotides encoding such fusion proteins. Examples of fusion protein technology can be found in International Patent Application Nos. WO94/29458 and WO94/22914.

Another aspect of the invention relates to a method for inducing an immunological response in a mammal which comprises inoculating the mammal with a polypeptide of the present invention, adequate to produce antibody and/or T cell immune response to protect said animal from the Diseases hereinbefore mentioned, amongst others. Yet another aspect of the invention relates to a method of inducing immunological response in a mammal which comprises, delivering a polypeptide of the present invention *via* a vector directing expression of the polynucleotide and coding for the polypeptide *in vivo* in order to induce such an immunological response to produce antibody to protect said animal from diseases.

Polypeptides of the present invention are responsible for many biological functions, including many disease states, in particular the Diseases hereinbefore mentioned. It is therefore desirous to devise screening methods to identify compounds which stimulate or which inhibit the function of the polypeptide. Accordingly, in a further aspect, the present invention provides for a method of screening compounds to identify those which stimulate or which inhibit the function of the polypeptide. In general, agonists or antagonists may be employed for therapeutic and prophylactic purposes for such Diseases as hereinbefore mentioned. Compounds may be identified from a variety

of sources, for example, cells, cell-free preparations, chemical libraries, and natural product mixtures. Such agonists, antagonists or inhibitors so-identified may be natural or modified substrates, ligands, receptors, enzymes, etc., as the case may be, of the polypeptide; or may be structural or functional mimetics thereof (see Coligan *et al.*, Current Protocols in Immunology 1(2):Chapter 5 (1991)).

The screening method may simply measure the binding of a candidate compound to the polypeptide, or to cells or membranes bearing the polypeptide, or a fusion protein thereof by means of a label directly or indirectly associated with the candidate compound. Alternatively, the screening method may involve competition with a labeled competitor. Further, these screening methods may test whether the candidate compound results in a signal generated by activation or inhibition of the polypeptide, using detection systems appropriate to the cells bearing the polypeptide. Inhibitors of activation are generally assayed in the presence of a known agonist and the effect on activation by the agonist by the presence of the candidate compound is observed. Constitutively active polpypeptides may be employed in screening methods for inverse agonists or inhibitors, in the absence of an agonist or inhibitor, by testing whether the candidate compound results in inhibition of activation of the polypeptide. Further, the screening methods may simply comprise the steps of mixing a candidate compound with a solution containing a polypeptide of the present invention, to form a mixture, measuring CASB7439 activity in the mixture, and comparing the CASB7439 activity of the mixture to a standard. Fusion proteins, such as those made from Fc portion and CASB7439 polypeptide, as hereinbefore described, can also be used for high-throughput screening assays to identify antagonists for the polypeptide of the present invention (see D. Bennett *et al.*, J Mol Recognition, 8:52-58 (1995); and K. Johanson *et al.*, J Biol Chem, 270(16):9459-9471 (1995)).

The polynucleotides, polypeptides and antibodies to the polypeptide of the present invention may also be used to configure screening methods for detecting the effect of added compounds on the production of mRNA and polypeptide in cells. For example, an ELISA assay may be constructed for measuring secreted or cell associated levels of polypeptide using monoclonal and polyclonal antibodies by standard methods known in the art. This can be used to discover agents which may inhibit or enhance the production of polypeptide (also called antagonist or agonist, respectively) from suitably manipulated cells or tissues.

The polypeptide may be used to identify membrane bound or soluble receptors, if any, through standard receptor binding techniques known in the art. These include, but are not limited to, ligand binding and crosslinking assays in which the polypeptide is labeled with a radioactive isotope (for instance, ^{125}I), chemically modified (for instance, biotinylated), or fused to a peptide sequence suitable for detection or purification, and incubated with a source of the putative

receptor (cells, cell membranes, cell supernatants, tissue extracts, bodily fluids). Other methods include biophysical techniques such as surface plasmon resonance and spectroscopy. These screening methods may also be used to identify agonists and antagonists of the polypeptide which compete with the binding of the polypeptide to its receptors, if any. Standard methods for conducting such assays are well understood in the art.

Examples of potential polypeptide antagonists include antibodies or, in some cases, oligonucleotides or proteins which are closely related to the ligands, substrates, receptors, enzymes, etc., as the case may be, of the polypeptide, e.g., a fragment of the ligands, substrates, receptors, enzymes, etc.; or small molecules which bind to the polypetide of the present invention but do not elicit a response, so that the activity of the polypeptide is prevented.

Thus, in another aspect, the present invention relates to a screening kit for identifying agonists, antagonists, ligands, receptors, substrates, enzymes, etc. for polypeptides of the present invention; or compounds which decrease or enhance the production of such polypeptides, which comprises:

(a) a polypeptide of the present invention;

(b) a recombinant cell expressing a polypeptide of the present invention;

(c) a cell membrane expressing a polypeptide of the present invention; or

(d) antibody to a polypeptide of the present invention;

which polypeptide is preferably that of SEQ ID NO:2 or SEQ ID NO:3.

It will be appreciated that in any such kit, (a), (b), (c) or (d) may comprise a substantial component.

It will be readily appreciated by the skilled artisan that a polypeptide of the present invention may also be used in a method for the structure-based design of an agonist, antagonist or inhibitor of the polypeptide, by:

(a) determining in the first instance the three-dimensional structure of the polypeptide;

(b) deducing the three-dimensional structure for the likely reactive or binding site(s) of an agonist, antagonist or inhibitor;

(c) synthesing candidate compounds that are predicted to bind to or react with the deduced binding or reactive site; and

(d) testing whether the candidate compounds are indeed agonists, antagonists or inhibitors.

It will be further appreciated that this will normally be an interative process.

In a further aspect, the present invention provides methods of treating abnormal conditions such as, for instance, cancer and autoimmune diseases, related to either an excess of, or an under-expression of, CASB7439 polypeptide activity.

If the activity of the polypeptide is in excess, several approaches are available. One approach comprises administering to a subject in need thereof an inhibitor compound (antagonist) as hereinabove described, optionally in combination with a pharmaceutically acceptable carrier, in an amount effective to inhibit the function of the polypeptide, such as, for example, by blocking the binding of ligands, substrates, receptors, enzymes, etc., or by inhibiting a second signal, and thereby alleviating the abnormal condition. In another approach, soluble forms of the polypeptides still capable of binding the ligand, substrate, enzymes, receptors, etc. in competition with endogenous polypeptide may be administered. Typical examples of such competitors include fragments of the CASB7439 polypeptide.

In still another approach, expression of the gene encoding endogenous CASB7439 polypeptide can be inhibited using expression blocking techniques. Known such techniques involve the use of antisense sequences, either internally generated or separately administered (see, for example, O'Connor, J Neurochem (1991) 56:560 in Oligodeoxynucleotides as Antisense Inhibitors of Gene Expression, CRC Press, Boca Raton, FL (1988)). Alternatively, oligonucleotides which form triple helices with the gene can be supplied (see, for example, Lee *et al.*, Nucleic Acids Res (1979) 6:3073; Cooney *et al.*, Science (1988) 241:456; Dervan *et al.*, Science (1991) 251:1360). These oligomers can be administered *per se* or the relevant oligomers can be expressed *in vivo*.

For treating abnormal conditions related to an under-expression of CASB7439 and its activity, several approaches are also available. One approach comprises administering to a subject a therapeutically effective amount of a compound which activates a polypeptide of the present invention, i.e., an agonist as described above, in combination with a pharmaceutically acceptable carrier, to thereby alleviate the abnormal condition. Alternatively, gene therapy may be employed to effect the endogenous production of CASB7439 by the relevant cells in the subject. For example, a polynucleotide of the invention may be engineered for expression in a replication defective retroviral vector, as discussed above. The retroviral expression construct may then be isolated and introduced into a packaging cell transduced with a retroviral plasmid vector containing RNA encoding a polypeptide of the present invention such that the packaging cell now produces infectious viral particles containing the gene of interest. These producer cells may be administered to a subject for engineering cells *in vivo* and expression of the polypeptide *in vivo*. For an overview of gene therapy, see Chapter 20, Gene Therapy and other Molecular Genetic-based Therapeutic Approaches, (and references cited therein) in Human Molecular Genetics, T Strachan and A P Read, BIOS Scientific Publishers Ltd (1996). Another approach is to administer a therapeutic amount of a polypeptide of the present invention in combination with a suitable pharmaceutical carrier.

In a further aspect, the present invention provides for pharmaceutical compositions comprising a therapeutically effective amount of a polypeptide, such as the soluble form of a

polypeptide of the present invention, agonist/antagonist peptide or small molecule compound, in combination with a pharmaceutically acceptable carrier or excipient. Such carriers include, but are not limited to, saline, buffered saline, dextrose, water, glycerol, ethanol, and combinations thereof. The invention further relates to pharmaceutical packs and kits comprising one or more containers filled with one or more of the ingredients of the aforementioned compositions of the invention. Polypeptides and other compounds of the present invention may be employed alone or in conjunction with other compounds, such as therapeutic compounds.

The composition will be adapted to the route of administration, for instance by a systemic or an oral route. Preferred forms of systemic administration include injection, typically by intravenous injection. Other injection routes, such as subcutaneous, intramuscular, or intraperitoneal, can be used. Alternative means for systemic administration include transmucosal and transdermal administration using penetrants such as bile salts or fusidic acids or other detergents. In addition, if a polypeptide or other compounds of the present invention can be formulated in an enteric or an encapsulated formulation, oral administration may also be possible. Administration of these compounds may also be topical and/or localized, in the form of salves, pastes, gels, and the like.

The dosage range required depends on the choice of peptide or other compounds of the present invention, the route of administration, the nature of the formulation, the nature of the subject's condition, and the judgment of the attending practitioner. Suitable dosages, however, are in the range of 0.1-100 μg/kg of subject. Wide variations in the needed dosage, however, are to be expected in view of the variety of compounds available and the differing efficiencies of various routes of administration. For example, oral administration would be expected to require higher dosages than administration by intravenous injection. Variations in these dosage levels can be adjusted using standard empirical routines for optimization, as is well understood in the art.

Polypeptides used in treatment can also be generated endogenously in the subject, in treatment modalities often referred to as "gene therapy" as described above. Thus, for example, cells from a subject may be engineered with a polynucleotide, such as a DNA or RNA, to encode a polypeptide *ex vivo*, and for example, by the use of a retroviral plasmid vector. The cells are then introduced into the subject.

Polynucleotide and polypeptide sequences form a valuable information resource with which to identify further sequences of similar homology. This is most easily facilitated by storing the sequence in a computer readable medium and then using the stored data to search a sequence database using well known searching tools, such as GCC. Accordingly, in a further aspect, the present invention provides for a computer readable medium having stored thereon a polynucleotide comprising the sequence of SEQ ID NO:1 and/or a polypeptide sequence encoded thereby.

The following definitions are provided to facilitate understanding of certain terms used frequently hereinbefore.

"Antibodies" as used herein includes polyclonal and monoclonal antibodies, chimeric, single chain, and humanized antibodies, as well as Fab fragments, including the products of an Fab or other immunoglobulin expression library.

"Isolated" means altered "by the hand of man" from the natural state. If an "isolated" composition or substance occurs in nature, it has been changed or removed from its original environment, or both. For example, a polynucleotide or a polypeptide naturally present in a living animal is not "isolated," but the same polynucleotide or polypeptide separated from the coexisting materials of its natural state is "isolated", as the term is employed herein.

"Polynucleotide" generally refers to any polyribonucleotide or polydeoxribonucleotide, which may be unmodified RNA or DNA or modified RNA or DNA. "Polynucleotides" include, without limitation, single- and double-stranded DNA, DNA that is a mixture of single- and double-stranded regions, single- and double-stranded RNA, and RNA that is mixture of single- and double-stranded regions, hybrid molecules comprising DNA and RNA that may be single-stranded or, more typically, double-stranded or a mixture of single- and double-stranded regions. In addition, "polynucleotide" refers to triple-stranded regions comprising RNA or DNA or both RNA and DNA. The term "polynucleotide" also includes DNAs or RNAs containing one or more modified bases and DNAs or RNAs with backbones modified for stability or for other reasons. "Modified" bases include, for example, tritylated bases and unusual bases such as inosine. A variety of modifications may be made to DNA and RNA; thus, "polynucleotide" embraces chemically, enzymatically or metabolically modified forms of polynucleotides as typically found in nature, as well as the chemical forms of DNA and RNA characteristic of viruses and cells. "Polynucleotide" also embraces relatively short polynucleotides, often referred to as oligonucleotides.

"Polypeptide" refers to any peptide or protein comprising two or more amino acids joined to each other by peptide bonds or modified peptide bonds, i.e., peptide isosteres. "Polypeptide" refers to both short chains, commonly referred to as peptides, oligopeptides or oligomers, and to longer chains, generally referred to as proteins. Polypeptides may contain amino acids other than the 20 gene-encoded amino acids. "Polypeptides" include amino acid sequences modified either by natural processes, such as post-translational processing, or by chemical modification techniques which are well known in the art. Such modifications are well described in basic texts and in more detailed monographs, as well as in a voluminous research literature. Modifications may occur anywhere in a polypeptide, including the peptide backbone, the amino acid side-chains and the amino or carboxyl termini. It will be appreciated that the same type of modification may be present to the same or varying degrees at several sites in a given polypeptide. Also, a given

polypeptide may contain many types of modifications. Polypeptides may be branched as a result of ubiquitination, and they may be cyclic, with or without branching. Cyclic, branched and branched cyclic polypeptides may result from post-translation natural processes or may be made by synthetic methods. Modifications include acetylation, acylation, ADP-ribosylation, amidation, covalent attachment of flavin, covalent attachment of a heme moiety, covalent attachment of a nucleotide or nucleotide derivative, covalent attachment of a lipid or lipid derivative, covalent attachment of phosphotidylinositol, cross-linking, cyclization, disulfide bond formation, demethylation, formation of covalent cross-links, formation of cystine, formation of pyroglutamate, formylation, gamma-carboxylation, glycosylation, GPI anchor formation, hydroxylation, iodination, methylation, myristoylation, oxidation, proteolytic processing, phosphorylation, prenylation, racemization, selenoylation, sulfation, transfer-RNA mediated addition of amino acids to proteins such as arginylation, and ubiquitination (see, for instance, Proteins - Structure and Molecular Properties, 2nd Ed., T. E. Creighton, W. H. Freeman and Company, New York, 1993; Wold, F., Post-translational Protein Modifications: Perspectives and Prospects, pgs. 1-12 in Post-translational Covalent Modification of Proteins, B. C. Johnson, Ed., Academic Press, New York, 1983; Seifter *et al.*, "Analysis for protein modifications and nonprotein cofactors", Meth Enzymol (1990) 182:626-646 and Rattan *et al.*, "Protein Synthesis: Post-translational Modifications and Aging", Ann NY Acad Sci (1992) 663:48-62).

"Variant" refers to a polynucleotide or polypeptide that differs from a reference polynucleotide or polypeptide, but retains essential properties. A typical variant of a polynucleotide differs in nucleotide sequence from another, reference polynucleotide. Changes in the nucleotide sequence of the variant may or may not alter the amino acid sequence of a polypeptide encoded by the reference polynucleotide. Nucleotide changes may result in amino acid substitutions, additions, deletions, fusions and truncations in the polypeptide encoded by the reference sequence, as discussed below. A typical variant of a polypeptide differs in amino acid sequence from another, reference polypeptide. Generally, differences are limited so that the sequences of the reference polypeptide and the variant are closely similar overall and, in many regions, identical. A variant and reference polypeptide may differ in amino acid sequence by one or more substitutions, additions, deletions in any combination. A substituted or inserted amino acid residue may or may not be one encoded by the genetic code. A variant of a polynucleotide or polypeptide may be a naturally occurring such as an allelic variant, or it may be a variant that is not known to occur naturally. Non-naturally occurring variants of polynucleotides and polypeptides may be made by mutagenesis techniques or by direct synthesis.

"Identity," as known in the art, is a relationship between two or more polypeptide sequences or two or more polynucleotide sequences, as determined by comparing the sequences. In the art, "identity" also means the degree of sequence relatedness between polypeptide or polynucleotide sequences,

as the case may be, as determined by the match between strings of such sequences. "Identity" and "similarity" can be readily calculated by known methods, including but not limited to those described in (Computational Molecular Biology, Lesk, A.M., ed., Oxford University Press, New York, 1988; Biocomputing: Informatics and Genome Projects, Smith, D.W., ed., Academic Press, New York, 1993; Computer Analysis of Sequence Data, Part I, Griffin, A.M., and Griffin, H.G., eds., Humana Press, New Jersey, 1994; Sequence Analysis in Molecular Biology, von Heinje, G., Academic Press, 1987; and Sequence Analysis Primer, Gribskov, M. and Devereux, J., eds., M Stockton Press, New York, 1991; and Carillo, H., and Lipman, D., SIAM *J. Applied Math., 48:* 1073 (1988). Preferred methods to determine identity are designed to give the largest match between the sequences tested. Methods to determine identity and similarity are codified in publicly available computer programs. Preferred computer program methods to determine identity and similarity between two sequences include, but are not limited to, the GCG program package (Devereux, J., et al., Nucleic Acids Research *12(1):* 387 (1984)), BLASTP, BLASTN, and FASTA (Atschul, S.F. et al., J. Molec. Biol. *215:* 403-410 (1990). The BLAST X program is publicly available from NCBI and other sources (BLAST Manual, Altschul, S., *et al.*, NCBI NLM NIH Bethesda, MD 20894; Altschul, S., *et al.*, J. Mol. Biol. *215*: 403-410 (1990). The well known Smith Waterman algorithm may also be used to determine identity.

Preferred parameters for polypeptide sequence comparison include the following:

1) Algorithm: Needleman and Wunsch, J. Mol Biol. 48: 443-453 (1970)

Comparison matrix: BLOSSUM62 from Hentikoff and Hentikoff, Proc. Natl. Acad. Sci. USA. 89:10915-10919 (1992)

Gap Penalty: 12

Gap Length Penalty: 4

A program useful with these parameters is publicly available as the "gap" program from Genetics Computer Group, Madison WI. The aforementioned parameters are the default parameters for polypeptide comparisons (along with no penalty for end gaps).

Preferred parameters for polynucleotide comparison include the following:

1) Algorithm: Needleman and Wunsch, J. Mol Biol. 48: 443-453 (1970)

Comparison matrix: matches = +10, mismatch = 0

Gap Penalty: 50

Gap Length Penalty: 3

A program useful with these parameters is publicly available as the "gap" program from Genetics Computer Group, Madison WI. The aforementioned parameters are the default parameters for polynucleotide comparisons.

By way of example, a polynucleotide sequence of the present invention may be identical to the reference sequence of SEQ ID NO:1, that is be 100% identical, or it may include up to a certain

integer number of nucleotide alterations as compared to the reference sequence. Such alterations are selected from the group consisting of at least one nucleotide deletion, substitution, including transition and transversion, or insertion, and wherein said alterations may occur at the 5' or 3' terminal positions of the reference nucleotide sequence or anywhere between those terminal positions, interspersed either individually among the nucleotides in the reference sequence or in one or more contiguous groups within the reference sequence. The number of nucleotide alterations is determined by multiplying the total number of nucleotides in SEQ ID NO:1 by the numerical percent of the respective percent identity(divided by 100) and subtracting that product from said total number of nucleotides in SEQ ID NO:1, or:

$$n_n \leq x_n - (x_n \bullet y),$$

wherein n_n is the number of nucleotide alterations, x_n is the total number of nucleotides in SEQ ID NO:1, and y is, for instance, 0.70 for 70%, 0.80 for 80%, 0.85 for 85%, 0.90 for 90%, 0.95 for 95%,etc., and wherein any non-integer product of x_n and y is rounded down to the nearest integer prior to subtracting it from x_n. Alterations of a polynucleotide sequence encoding the polypeptide of SEQ ID NO:2 may create nonsense, missense or frameshift mutations in this coding sequence and thereby alter the polypeptide encoded by the polynucleotide following such alterations.

Similarly, a polypeptide sequence of the present invention may be identical to the reference sequence of SEQ ID NO:2, that is be 100% identical, or it may include up to a certain integer number of amino acid alterations as compared to the reference sequence such that the % identity is less than 100%. Such alterations are selected from the group consisting of at least one amino acid deletion, substitution, including conservative and non-conservative substitution, or insertion, and wherein said alterations may occur at the amino- or carboxy-terminal positions of the reference polypeptide sequence or anywhere between those terminal positions, interspersed either individually among the amino acids in the reference sequence or in one or more contiguous groups within the reference sequence. The number of amino acid alterations for a given % identity is determined by multiplying the total number of amino acids in SEQ ID NO:2 by the numerical percent of the respective percent identity(divided by 100) and then subtracting that product from said total number of amino acids in SEQ ID NO:2, or:

$$n_a \leq x_a - (x_a \bullet y),$$

wherein n_a is the number of amino acid alterations, x_a is the total number of amino acids in SEQ ID NO:2, and y is, for instance 0.70 for 70%, 0.80 for 80%, 0.85 for 85% etc., and wherein any non-integer product of x_a and y is rounded down to the nearest integer prior to subtracting it from x_a.

"Fusion protein" refers to a protein encoded by two, often unrelated, fused genes or fragments thereof. In one example, EP-A-0 464 discloses fusion proteins comprising various portions of constant region of immunoglobulin molecules together with another human protein or part thereof. In many cases, employing an immunoglobulin Fc region as a part of a fusion protein is advantageous for use in therapy and diagnosis resulting in, for example, improved pharmacokinetic properties [see, e.g., EP-A 0232 262]. On the other hand, for some uses it would be desirable to be able to delete the Fc part after the fusion protein has been expressed, detected and purified.

All publications, including but not limited to patents and patent applications, cited in this specification are herein incorporated by reference as if each individual publication were specifically and individually indicated to be incorporated by reference herein as though fully set forth.

Examples

Subtractive cDNA cloning of colon tumour-associated antigen (TAA) candidates.

Subtractive cDNA libraries are produced using standard technologies. Briefly, total RNA is extracted from frozen (-70°C) tumour and matched normal colon samples using the TriPure reagent and protocol (Boehringer). Target RNA is prepared by pooling total RNA from three tumour samples (30 µg each). Driver RNA is prepared by pooling total RNA from three matched normal colon samples (10 µg each) and total RNA from seven normal tissues other than colon(brain, heart, kidney, liver, bladder, skin, spleen; 10 µg each). Total RNA from non-colon normal tissues is purchased from InVitrogen.

Messenger RNA is purified from total RNA using oligo-dT magnetic bead technology (Dynal) and quantified by spectrofluorimetry (BioRad).

Target and driver mRNA are reverse transcribed into cDNA using one of two strategies: 1) Target sequences for PCR oligonucleotides are introduced onto the ends of the newly synthesised cDNA during reverse transcription using the template switching capability of reverse transcriptase (ClonTech SMART PCR cDNA synthesis kit). 2) Alternatively, the target and driver mRNA are reverse transcribed into cDNA using an oligo-dT primer and converted to double-strand cDNA; the cDNA is cleaved with RsaI and linkers for PCR amplification are ligated onto the extremities of the cDNA fragments.

In both cases, target and driver cDNA are amplified by long range PCR (ClonTech SMART PCR Synthesis Kit and Advantage PCR Polymerase Mix) and used as starting material for

subtractive cloning. For amplification, cycling conditions and optimisation of the number of PCR cycles are as described in the Advantage PCR protocol.

Two subtractive cloning strategies are used: ClonTech PCR SELECT (see ClonTech kit protocol and N. Gurskaya *et al.* 1996. Analytical Biochemistry: 240, 90) and cRDA(M. Hubank and D. Schatz. 1994. Nucleic Acids Research: 22, 5640). When the PCR SELECT protocol is used, the primary PCR SELECT subtraction products are submitted to a supplementary round of cRDA subtraction. When the cRDA protocol is used, two consecutive cycles of cRDA subtraction are performed. In each case the products of both cycles of subtraction are cloned into pCR-TOPO (InVitrogen) and transformed into E. coli to produce a subtracted cDNA plasmid library.

An alternative strategy is also followed: subtraction of normal colon sequences and sequences from non-colon normal tissues are subtracted in separate hybridizations. In this case, target and driver RNA are assembled for the first subtraction as above with the exception that non-colon RNA is left out of the driver pool and amounts of normal colon are increased to 10 μg. Preparation of target and driver cDNA and subtractive hybridization are performed as described above. A second subtraction is then performed on the products of the first subtraction, but the driver is now composed of a pool of normal colon and normal non-colon mRNA from the seven normal tissues.

Differential Screening of cDNA arrays.

Identification of tumour-associated genes in the subtracted cDNA library is accomplished by differential screening.

Total bacterial DNA is extracted from 100 μl over-night cultures. Bacteria are lysed with guanidium isothiocyantate and the bacterial DNA is affinity purified using magnetic glass (Boehringer). Plasmid inserts are recovered from the bacterial DNA by Advantage PCR amplification (Clontech). The PCR products are dotted onto two nylon membranes to produce high density cDNA arrays using the Biomek 96 HDRT tool (Beekman). The spotted cDNA is covalently linked to the membrane by UV irradiation. The first membrane is hybridised with a mixed cDNA probe prepared from the tumour of a single patient. The second membrane is hybridised with an equivalent amount of mixed cDNA probe prepared from normal colon of the same patient. The probe cDNA is prepared by PCR amplification as described above and is labelled using the AlkPhos Direct System (Amersham). Hybridisation conditions and stringency washes are as described in the AlkPhos Direct kit. Hybridized probe is detected by chemiluminescence. Hybridisation intensities for each cDNA fragment on both blots are measured by film densitometry or direct measurement (BioRad Fluor-S Max). The ratio of the tumour to normal hybridisation intensities (T/N) is calculated for each gene to evaluate the degree of over-expression in the tumour. Genes which are significantly over-expressed in colon tumours are followed-up. Significance is arbitrarily defined as one standard

deviation of the T/N frequency distribution. Differential screening experiments are repeated using RNA from multiple patient donors (>18) to estimate the frequency of over-expressing tumours in the patient population.

In addition, the DNA arrays are hybridised with mixed cDNA probes from normal tissues other than colon (see list above) to determine the level of expression of the candidate gene in these tissues.

Real-time RT-PCR analysis

Real-time RT-PCR (U. Gibson. 1996. Genome Research: 6,996) is used to compare mRNA transcript abundance of the candidate antigen in matched tumour and normal colon tissues from multiple patients. In addition, mRNA levels of the candidate gene in a panel of normal tissues are evaluated by this approach.

Total RNA from normal and tumour colon is extracted from snap frozen biopsies using TriPure reagent (Boehringer). Total RNA from normal tissues is purchased from InVitrogen or is extracted from snap frozen biopsies using TriPure reagent. Poly A+ mRNA is purified from total RNA after DNAase treatment using oligo-dT magnetic beads (Dynal). Quantification of the mRNA is performed by spectrofluorimetry (VersaFluor, BioRad) using SybrII dye (Molecular Probes). Primers for real-time PCR amplification are designed with the Perkin-Elmer Primer Express software using default options for TaqMan amplification conditions.

Real-time reactions are assembled according to standard PCR protocols using 2 ng of purified mRNA for each reaction. Sybrl dye (Molecular Probes) is added at a final dilution of 1/75000 for real-time detection. Amplification (40 cycles) and real-time detection is performed in a Perkin-Elmer Biosystems PE7700 system using conventional instrument settings. Ct values are calculated using the PE7700 Sequence Detector software. Two Ct values are obtained for each patient sample: the tumour Ct (CtT) and the matched normal colon Ct (CtN). Ct values obtained by real-time PCR are log-linearly related to the copy number of the target template. As the efficiency of PCR amplification under the prevailing experimental conditions is close to the theoretical amplification efficiency, 2 (CtN-CtT) is an estimate of the relative transcript levels in the two tissues (i.e. fold mRNA over-expression in tumor). Real-time PCR reactions are performed on biopsies from 12 patients. The level of mRNA over-expression is calculated as described for each patient. average level of mRNA over-expression for the candidate antigen and the proportion of patients over-expressing the candidate antigen is then calculated from this data set. The individual values are standardised with respect to actin in the same sample (ratio), as seen in Figure 1. A value of 1 thus corresponds to the same level of actin expression. The results are shown in a logarithmic scale.

A series of 36 normal tissue samples, representing 28 different tissues, were also tested by the same procedure. Ct values for the candidate antigen were compared to those of actin obtained with the same tissue sample. The results, standardized with respect to actin, are shown in Figure 2.

Real-time PCR results in colon cancer/normal colon sample

Summary

Patients over-expressing CASB7439 in colon tumours (%)	Average level of over-expression in colon tumours (fold)
11/12 (90%)	2 200

Conclusion: CASB7439 is overexpressed in 90% of colon cancer samples with respect to the adjacent normal colon, at an average rate of nearly 2 200 fold. The expression in normal tissues is restricted to prostate and stomach.

Northern-Southern blot analysis

Limited amounts of mixed tumor and matched normal colon cDNA are amplified by Advantage PCR (see above). Messenger RNA from multiple normal tissues is also amplified using the same procedure. The amplified cDNA (1 µg) is electrophoresed on a 1.2% agarose gel and transferred onto a nylon membrane. The membrane is hybridised (AlkPhos Direct System) with a probe prepared using a fragment of the candidate TAA cDNA. Northern-Southern analysis provides information on transcript size, presence of splice variants and transcript abundance in tumour and normal tissues. Northern-Southern Blot analysis is used in place of the more familiar Northern Blot analysis as most tissue samples are too small to yield enough RNA for the latter procedure.

SEQUENCE INFORMATION

SEQ ID NO:1

GTACCTTGCTTTGGGGGCGCACTAAGTACCTGCCGGGAGCAGGGGGCGCACCGGGAACTCGCAGATTTCGCCAGTTGGGC
GCACTGGGGATCTGTGGACTGCGTCCGGGGGATGGGCTAGGGGCGACATGCGCACGCTTTGGGCCTTACAGAATGTGATCG
CGCCGAGGGGGAGGGCCGAAGCGTGGCGGGAGGGCGAGGCGAAGGAAGGAGGGCGTGAGAAAGGCGACGGCGGCGGCGCG
GAGGAGGGTTATCTATACATTTAAAAACCAGCCGCCTGCGCCGCGCCTGCGGAGACCTGGGAGAGTCCGGCCCCACGCCC
GGGACACGAGCGTCCCACGCTCCCTGGCGCGTACGGCCTGCCACCACTAGGCCTCCTATCCCCGGGCTCCAGACGACCTA
GGACGCGTGCCCTGGGGAGTTGCCTGGCGGCGCCGTGCCAGAAGCCCCCTTGGGGCGCCACAGTTTTCCCCGTCGCCTCC
GGTTCCTCTGCCTGCACCTTCCTGCGGCGCGCCGGGACCTGGAGCGGCGGGTGGATGCAGGCGCGatggacggcggcac
actgcccaggtccgcgcccccctgcgcccccgtcccctgtcggctgcgctgcccyggyagacccgcgtccccggaactgt
tgcgctgcagccggcggcggcgaccggcaacgcagagaccggaggcggcgcagcggccgtagcgcggcgcaatgagcgc
gagcgcaaccgcgtgaagctggcgaacttgggcttccaggcgctgcggcagcacgtgccgcacgqccqqcgccagcaagaa
gctgagcaaggtggagacgctgcgctcagccgtggagtacatccgcgcgctgcagcgcctgctggccgagcacgacgccg
cgcgcaacgcgctggcgggagggctgaggccgcaggccgtgcggccgtctgcgccccgcggcgccgccagggaccaccccg
gtcgccgcctcgccctcccgcgcttcttcgtccccgggccgcggggggcagctcggagccccggctcccccgcgttccgccta
ctcgtcggacgacagcggctgcgaagggcgcgctgagtcctgcggagcgcgagctactcgacttctccagctggttagggg
gctactgaGCGCCCTCGACCTAATAAGCCTCAAGCCCCGGAAACCCGAGCGAACGGGCCGGCGCGCTTCATCGCCGGGGA
AGCCCGCCAAGGTGGACCGGGCCCGCGCTCCGCCCCCAGCGAGCCGGGGACCCACCCACCACCACCCCCCGCACCGCCGACGC
CGCCTCGTTCGTCCGGCCCAGCCTGACCAATGCCGCGGTGGAAACGGGCTTGGAGCTGGCCCCATAAGGGCTGGCGGCTT
CCTCCGACGCCGCCCCTCCCCACAGCTTCTCGACTGCAGTGGGGCGGGGGGCACCAACACTTGGAGATTTTTCCGGAGGG
GAGAGGATTTTCTAAGGGCACAGAGAATCCATTTTCTACACATTAACTTGAGCTGCTGGAGGGACACTGCTGGCAAACGG
AGACCTATTTTTGTACAAAGAACCCTTGACCTGGGGCGTAATAAAGATGACCTGGACCCCTGCCCCCACTATCTGGAGTT
TTCCATGCTGGCCAAGATCTGGACACGAGCAGTCCCTGAGGGGCGGGGTCCCTGGCGTGAGGCCCCCGTGACAGCCCACC
CTGGGGTGGGTTTGTGGGCACTGCTGCTCTGCTAGGGAGAAGCCTGTGTGGGGCACACCTCTTCAAGGGAGCGTGAACTT
TATAAATAAATCAGTTCTGTTTAAAAAAAAAAAAAAAAAAAAACCGAGGGGGGGCCCGGAGCCAACAAA

SEQ ID NO:2

mdggllprsappappvpvgcaarrrpaspellrcsrrrrpataetgggaaavarrnererrnrvklvnlgfqalrqrvphg
gaskklskvetlrsaveyiralqrllaehdavrnalagglrpqavrpsaprgppgttpvaaspsraaasspgrggssepgs
prsavssddsqceqalspaerelldfsswlggy

SEQ ID NO:3

mcrkwilcalrkssplrknlqvlvppaplqsrscgegrrrrkppalmgpapspfpprhwsgwagrtrrrrrcggwwvgpr
lagggararstlagfpgdearrpvrsgfrglrlirsralsspltswrsrvarapqdsarlrsrcrptsrrnagsrapscp
rgpgtkkrgrarrrpgwslaargaqtaerpaasalpparcarrrarpagaaargctprlsaasppcsascwrrraaraaa
apgspsspaasrgcarahcaalrplrrlrslrwpvaaagcsatvpgtrvsagqrsrqgrgaqgartwavcrrpsrlhppar
srsrraagrcrqrnrrrrgklwrpkgasgtappgnspghas

SEQ ID NO:4

GGTAAACAGAACTGATTTATTTATAAAGTTCACGCTCCCTTGAAGAGGTGTGCCCCACACAGGCTTCTCCCTAGCAGAGC
AGCAGTGCCCACAAACCCACCCCAGGGTGGGCTGTCACGGGGGCCTCACGCCAGGGACCCCGCCCCTCAGGGACTGCTCG
TGTCCAGATCTTGGCCAGCATGGAAAACTCCAGATAGTGGGGGCAGGGGTCCAGGTCATCTTTATTACGCCCCAGGTCAA
GGGTTCTTTGTACAAAAATAGGTCTCCGTTTGCCAGCAGTGTCCCTCCAGCAGCTCAAGTTAATGTGTAGAAAATGGATT
CTCTGTGCCCTTAGAAAATCCTCTCCCCTCCGGAAAAATCTCCAAGTGTTGGTGCCCCCCGCCCCACTGCAGTCGAGAAG
CTGTGGGGAGGGGCGGCGTCGGAGGAAGCCGCAGCCCATTATGGGGCCAGCTCCAAGCCCGTTTCCACCGCGGCATTGGT
CAGGCTGGGCGGACGAACGAGGCGGCGTCGGCGGTGCGGGGGGTGGTGGGTGGGTCCCCGGCTCGCTGGGGGCGGAGCAG
CGGGCCGGTCCACCTGGCGGGCTCCCC

What is claimed is:

1. An isolated polynucleotide comprising a nucleotide sequence that has at least 70% identity to a nucleotide sequence of SEQ ID NO:1 over its entire length; or a nucleotide sequence complementary to said isolated polynucleotide.

2. The polynucleotide of claim 1 which is DNA or RNA.

3. The polynucleotide of claim 1 which is the polynucleotide of SEQ ID NO: 1.

4. An isolated polynucleotide comprising a polynucleotide that hybridizes under stringent condition to a nucleotide sequence having SEQ ID NO: 1 or a fragment thereof.

5. A polypeptide comprising amino acid sequence encoded by the polynucleotide obtained in claim 4.

6. A vaccine comprising an effective amount of the polynucleotide of claim 1 and a pharmaceutically acceptable carrier

7. A vaccine comprising an effective amount of the polypeptide of claim 5 and a pharmaceutically effective carrier.

8. Use of the vaccines as claimed in claims 6 and 7 in the immunoprophylaxis or therapy of disease.

Abstract

CASB7439 polypeptides and polynucleotides and methods for producing such polypeptides by recombinant techniques are disclosed. Also disclosed are methods for utilizing CASB7439 polypeptides and polynucleotides in diagnostics, prophylaxis and therapy, and diagnostic assays for such.

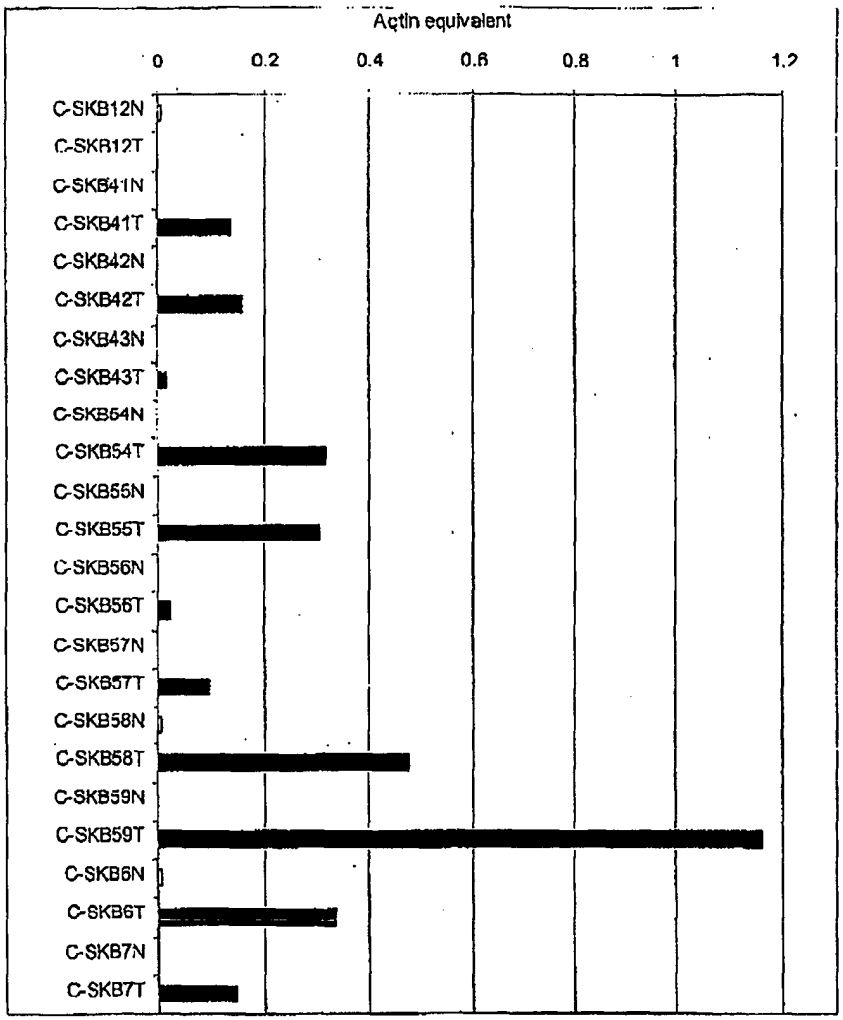


Figure 1: Real-time PCR data on matched normal and tumoral colon samples

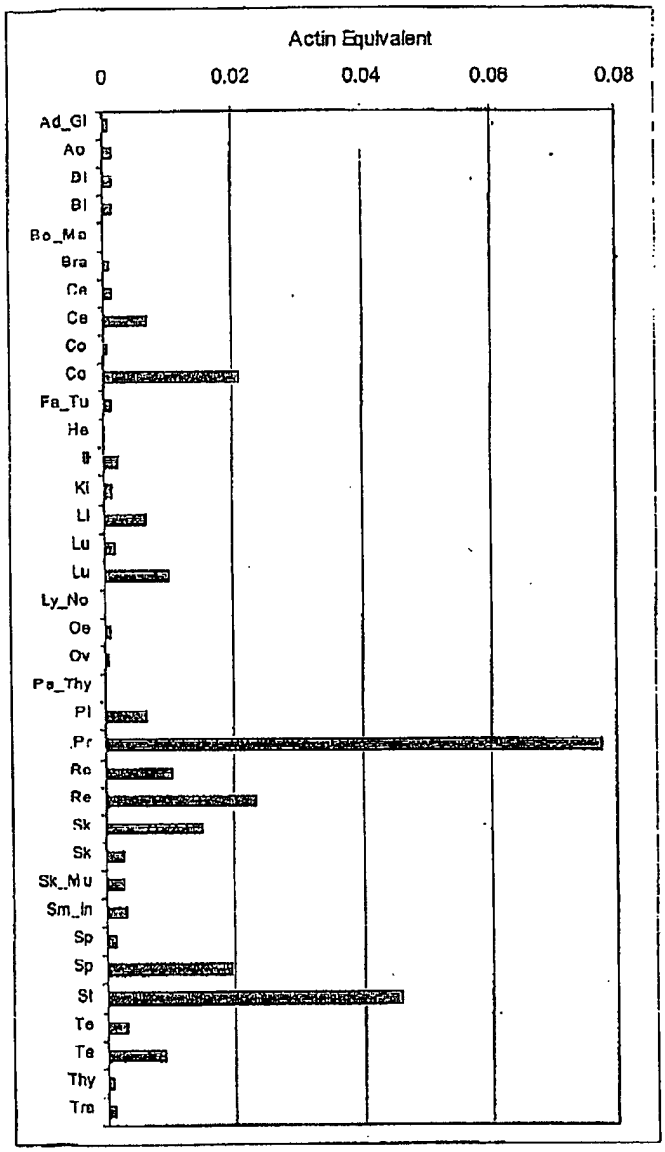


Figure 2 : Real-time PCR data of CASB7439 in normal tissues

Legend: Ad_Gl: adrenal gland; Ao: aorta; Bl: bladder, Bo_Ma: bone marrow; Bra: brain; Ce: cervix; Co: colon; Fa_Tu: fallopian tube; He: heart; Il: ileon; Ki: kidney; Li: liver; Lu: lung; Ly_No: lymph node; Oe: oesophagus; Ov: ovary; a_Thy: parathyroid gland; Pl: placenta; Pr: prostate; Re: rectum; Sk: skin; Sk_Mu: skeletal muscle; Sp: pleen; St: stomach; Te: testis; Thy: thyroid gland; Tr: trachea

The Patent Office
Concept House
Cardiff Road
Newport
South Wales
NP10 8QQ

PRIORITY DOCUMENT

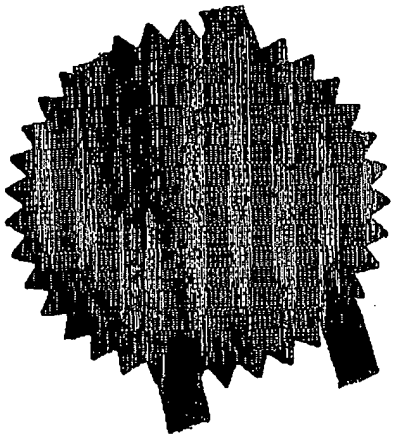
SUBMITTED OR TRANSMITTED IN COMPLIANCE WITH RULE 17.1(a) OR (b)

I, the undersigned, being an officer duly authorised in accordance with Section 74(1) and (4) of the Deregulation & Contracting Out Act 1994, to sign and issue certificates on behalf of the Comptroller-General, hereby certify that annexed hereto is a true copy of the documents as originally filed in connection with the patent application identified therein.

In accordance with the Patents (Companies Re-registration) Rules 1982, if a company named in this certificate and any accompanying documents has re-registered under the Companies Act 1980 with the same name as that with which it was registered immediately before re-registration save for the substitution as, or inclusion as, the last part of the name of the words "public limited company" or their equivalents in Welsh, references to the name of the company in this certificate and any accompanying documents shall be treated as references to the name with which it is so re-registered.

In accordance with the rules, the words "public limited company" may be replaced by p.l.c., plc, P.L.C. or PLC.

Re-registration under the Companies Act does not constitute a new legal entity but merely subjects the company to certain additional company law rules.

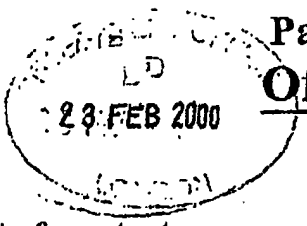


Signed A W Brewer.

Dated 31 May 2001

Patents Form 1/77

Patents Act 1977
(Rule 16)



The Patent Office

1/77

24FEB00 E515815-1 D02029
P01/7700 0.00-0004269.7

Request for grant of a patent

(See the notes on the back of this form. You can also get an explanatory leaflet from the Patent Office to help you fill in this form)

The Patent Office
Cardiff Road
Newport
Gwent NP9 1RH

1. Your reference — KLP/RSH/BC45300

2. Patent application number
(The Patent Office will fill in this part) — 23 FEB 2000 — **0004269.7**

3. Full name, address and postcode of the or of each applicant *(underline all surnames)* — SmithKline Beecham Biologicals s.a.
rue de l'Institut 89, B-1330 Rixensart, Belgium

Patents ADP number *(if you know it)*

If the applicant is a corporate body, give the country/state of its incorporation — Belgium — 6570956 00

4. Title of the invention — NOVEL COMPOUNDS

5. Name of your agent *(if you have one)* — CORPORATE INTELLECTUAL PROPERTY

"Address for service" in the United Kingdom to which all correspondence should be sent
(including the postcode) — SMITHKLINE BEECHAM PLC
TWO NEW HORIZONS COURT
BRENTFORD
MIDDLESEX TW8 9EP — 5800974 00

Patents ADP number *(if you know it)*

6. If you are declaring priority from one or more earlier patent applications, give the country and the date of filing of the or each of these earlier applications and *(if you know it)* the or each application number

Country	Priority application number *(if you know it)*	Date of filing *(day / month / year)*

7. If this application is divided or otherwise derived from an earlier UK application, give the number and the filing date of the earlier application

Number of earlier application	Date of filing *(day / month / year)*

8. Is a statement of inventorship and of right to grant of a patent required in support of this request? *(Answer yes if:*
a) any applicant named in part 3 is not an inventor, or
b) there is an inventor who is named as an applicant, or
c) any named applicant is a corporate body
See note (d)

Patents Form 1/77

9. Enter the number of sheets for any of the following items you are filing with this form. Do not count copies of the same document

Continuation sheets of this form	
Description	22
Claim(s)	1
Abstract	1
Drawings	

10. If you are also filing any of the following, state how many against each item.

Priority Documents	
Translations of priority documents	
Statement of inventorship and right to grant of a patent *(Patents Form 1/77)*	
Request for preliminary examination and search *(Patents Form 9/77)*	
Request for substantive examination *(Patents Form 10/77)*	
Any other documents *(please specify)*	

11. We request the grant of a patent on the basis of this application
Signature K L Privett Date 18-Feb-00

12. Name and daytime telephone number of person to contact in the United Kingdom: K L Privett 0181 9752585

Warning

After an application for a Patent has been filed, the Comptroller of the Patent Office will consider whether publication or communication of the invention should be prohibited or restricted under Section 22 of the Patents Act 1977. You will be informed if it is necessary to prohibit or restrict your invention in this way. Furthermore, if you live in the United Kingdom, Section 23 of the Patents Act 1977 stops you from applying for a patent abroad without first getting written permission unless an application has been filed at least six weeks beforehand in the United Kingdom for a patent for the same invention and either no direction prohibiting publication or communication has been given, or any such direction has been revoked.

Notes

a) *If you need help to fill in this form or you have any questions, please contact the Patent Office on 0645 500505*
b) *Write your answers in capital letters using black ink or you may type them.*
c) *If there is not enough space for all relevant details on any part of this form, please continue on a separate sheet of paper and write "see continuation sheet" in the relevant part(s). Any continuation sheet should be attached to this form.*
d) *If you have answered 'Yes' Patents Form 7/77 will need to be filed.*

Novel Compounds

Field of the Invention

This invention relates to newly identified polypeptides and polynucleotides encoding such polypeptides, to their use in diagnostics, prophylaxis and therapy and in identifying compounds which may be agonists, antagonists and /or inhibitors which are potentially useful in therapy, and to production of such polypeptides and polynucleotides.

Background of the Invention

The drug discovery process is currently undergoing a fundamental revolution as it embraces 'functional genomics', that is, high throughput genome- or gene-based biology. This approach is rapidly superceding earlier approaches based on 'positional cloning'. A phenotype, that is a biological function or genetic disease, would be identified and this would then be tracked back to the responsible gene, based on its genetic map position.

Functional genomics relies heavily on high-throughput DNA sequencing technologies and the various tools of bioinformatics to identify gene sequences of potential interest from the many molecular biology databases now available. cDNA libraries enriched for genes of relevance to a particular tissue or physiological situation can be constructed using recently developed subtractive cloning strategies. Construction of subtractive libraries reduces cDNA sequence complexity, focuses experimental resources on relevant genes and thus greatly accelerates the gene identification process. There is a continuing need to identify and characterise further genes and their related polypeptides/proteins, as targets for drug discovery.

High throughput genome- or gene-based biology allows new approaches to the identification and cloning of target genes for useful immune responses for the prevention and vaccine therapy of diseases such as cancer and autoimmunity.

Summary of the Invention

The present invention relates to CASB7439, in particular CASB7439 polypeptides and CASB7439 polynucleotides, recombinant materials and methods for their production. In another aspect, the invention relates to methods for using such polypeptides and polynucleotides, including the treatment of cancer and autoimmune diseases, hereinafter referred to as "the Diseases", amongst others. In a further aspect, the invention relates to methods for identifying agonists and antagonists/inhibitors using the materials provided by the invention, and treating conditions associated with CASB7439 imbalance with the identified compounds In a still further aspect, the

invention relates to diagnostic assays for detecting diseases associated with inappropriate CASB7439 activity or levels.

Description of the Invention

In a first aspect, the present invention relates to CASB7439 polypeptides. Such peptides include isolated polypeptides encoded by a polynucleotide comprising the sequence contained in SEQ ID NO:1.

Polypeptides and polynucleotides of the present invention are believed to be important immunogens for specific prophylactic or therapeutic immunization against tumors, because they are highly overexpressed in tumors and can thus be targeted by antigen-specific immune mechanisms leading to the destruction of the tumor cell. They can also be used to diagnose the occurrence of tumor cells. Furthermore, their inappropriate expression in certain circumstances can cause an induction of autoimmune; inappropriate immune responses, which could be corrected through appropriate vaccination using the same polypeptides or polynucleotides. In this respect the most important biological activities to our purpose are the antigenic and immunogenic activities of the polypeptide of the present invention. A polypeptide of the present invention may also exhibit at least one other biological activity of CASB7439, which could qualify it as a target for therapeutic or prophylactic intervention different from that linked to the immune response.

The polypeptides of the present invention may be in the form of the "mature" protein or may be a part of a larger protein such as a fusion protein. It is often advantageous to include an additional amino acid sequence which contains secretory or leader sequences, pro-sequences, sequences which aid in purification such as multiple histidine residues, or an additional sequence for stability during recombinant production. Furthermore, addition of exogenous polypeptide or lipid tail or polynucleotide sequences to increase the immunogenic potential of the final molecule is also considered.

The present invention also includes variants of the aforementioned polypeptides, that is polypeptides that vary from the referents by conservative amino acid substitutions, whereby a residue is substituted by another with like characteristics. Typical such substitutions are among Ala, Val, Leu and Ile; among Ser and Thr; among the acidic residues Asp and Glu; among Asn and Gln; and among the basic residues Lys and Arg; or aromatic residues Phe and Tyr. Particularly preferred are variants in which several, 5-10, 1-5, 1-3, 1-2 or 1 amino acids are substituted, deleted, or added in any combination.

Polypeptides of the present invention can be prepared in any suitable manner. Such polypeptides include isolated naturally occurring polypeptides, recombinantly produced

polypeptides, synthetically produced polypeptides, or polypeptides produced by a combination of these methods. Means for preparing such polypeptides are well understood in the art.

Further polynucleotides of the present invention include isolated polynucleotides comprising a nucleotide sequence which has at least 70% identity, preferably at least 80% identity, more preferably at least 90% identity, yet more preferably at least 95% identity, to SEQ ID NO:1 over the entire length of SEQ ID NO:1. In this regard, polynucleotides which have at least 97% identity are highly preferred, whilst those with at least 98-99% identiy are more highly preferred, and those with at least 99% identity are most highly preferred. Such polynucleotides include a polynucleotide comprising the polynucleotide of SEQ ID NO:1 as well as the polynucleotide of SEQ ID NO:1. Said polynucleotide can be inserted in a suitable plasmid or recombinant microrganism vector and used for immunization(see for example Wolff et. al., Science 247:1465-1468 (1990); Corr et. al., J. Exp. Med. 184:1555-1560 (1996); Doe et. al., Proc. Natl. Acad. Sci. 93:8578-8583 (1996)).

The invention also provides polynucleotides which are complementary to all the above described polynucleotides.

Preferred polypeptides and polynucleotides of the present invention are expected to have, *inter alia*, similar biological functions/properties to their homologous polypeptides and polynucleotides. Furthermore, preferred polypeptides and polynucleotides of the present invention have at least one CASB7439 activity.

Polynucleotides of the present invention may be obtained, using standard cloning and screening techniques, from a cDNA library derived from mRNA in cells of human colon cancer, using the expressed sequence tag (EST) analysis (Adams, M.D., *et al.* Science (1991) 252:1651-1656; Adams, M.D. *et al.*, Nature, (1992) *355*:632-634; Adams, M.D., *et al.*, Nature (1995) 377 Supp:3-174). Polynucleotides of the invention can also be obtained from natural sources such as genomic DNA libraries or can be synthesized using well known and commercially available techniques.

When polynucleotides of the present invention are used for the recombinant production of polypeptides of the present invention, the polynucleotide may include the coding sequence for the mature polypeptide, by itself; or the coding sequence for the mature polypeptide in reading frame with other coding sequences, such as those encoding a leader or secretory sequence, a pre-, or pro- or prepro- protein sequence, or other fusion peptide portions. For example, a marker sequence

which facilitates purification of the fused polypeptide can be encoded. In certain preferred embodiments of this aspect of the invention, the marker sequence is a hexa-histidine peptide, as provided in the pQE vector (Qiagen, Inc.) and described in Gentz *et al.*, Proc Natl Acad Sci USA (1989) 86:821-824, or is an HA tag. The polynucleotide may also contain non-coding 5' and 3' sequences, such as transcribed, non-translated sequences, splicing and polyadenylation signals, ribosome binding sites and sequences that stabilize mRNA.

Polynucleotides which are identical or sufficiently identical to a nucleotide sequence contained in SEQ ID NO:1, may be used as hybridization probes for cDNA and genomic DNA or as primers for a nucleic acid amplification (PCR) reaction, to isolate full-length cDNAs and genomic clones encoding polypeptides of the present invention and to isolate cDNA and genomic clones of other genes (including genes encoding homologs and orthologs from species other than human) that have a high sequence similarity to SEQ ID NO:1. Typically these nucleotide sequences are 70% identical, preferably 80% identical, more preferably 90% identical, most preferably 95% identical to that of the referent. The probes or primers will generally comprise at least 15 nucleotides, preferably, at least 30 nucleotides and may have at least 50 nucleotides. Particularly preferred probes will have between 30 and 50 nucleotides. In particular, polypeptides or polynucleotides derived from sequences from homologous animal origin could be used as immunogens to obtain a cross-reactive immune response to the human gene.

A polynucleotide encoding a polypeptide of the present invention, including homologs and orthologs from species other than human, may be obtained by a process which comprises the steps of screening an appropriate library under stringent hybridization conditions with a labeled probe having the sequence of SEQ ID NO: 1 or a fragment thereof; and isolating full-length cDNA and genomic clones containing said polynucleotide sequence. Such hybridization techniques are well known to the skilled artisan. Preferred stringent hybridization conditions include overnight incubation at 42°C in a solution comprising: 50% formamide, 5xSSC (150mM NaCl, 15mM trisodium citrate), 50 mM sodium phosphate (pH7.6), 5x Denhardt's solution, 10 % dextran sulfate, and 20 microgram/ml denatured, sheared salmon sperm DNA; followed by washing the filters in 0.1x SSC at about 65°C. Thus the present invention also includes polynucleotides obtainable by screening an appropriate library under stingent hybridization conditions with a labeled probe having the sequence of SEQ ID NO:1 or a fragment thereof.

The skilled artisan will appreciate that, in many cases, an isolated cDNA sequence will be incomplete, in that the region coding for the polypeptide is cut short at the 5' end of the cDNA. This is a consequence of reverse transcriptase, an enzyme with inherently low 'processivity' (a measure of the ability of the enzyme to remain attached to the template during the polymerisation

reaction), failing to complete a DNA copy of the mRNA template during 1st strand cDNA synthesis.

There are several methods available and well known to those skilled in the art to obtain full-length cDNAs, or extend short cDNAs, for example those based on the method of Rapid Amplification of cDNA ends (RACE) (see, for example, Frohman et al., PNAS USA 85, 8998-9002, 1988). Recent modifications of the technique, exemplified by the Marathon™ technology (Clontech Laboratories Inc.) for example, have significantly simplified the search for longer cDNAs. In the Marathon™ technology, cDNAs have been prepared from mRNA extracted from a chosen tissue and an 'adaptor' sequence ligated onto each end. Nucleic acid amplification (PCR) is then carried out to amplify the 'missing' 5' end of the cDNA using a combination of gene specific and adaptor specific oligonucleotide primers. The PCR reaction is then repeated using 'nested' primers, that is, primers designed to anneal within the amplified product (typically an adaptor specific primer that anneals further 3' in the adaptor sequence and a gene specific primer that anneals further 5' in the known gene sequence). The products of this reaction can then be analysed by DNA sequencing and a full-length cDNA constructed either by joining the product directly to the existing cDNA to give a complete sequence, or carrying out a separate full-length PCR using the new sequence information for the design of the 5' primer.

Recombinant polypeptides of the present invention may be prepared by processes well known in the art from genetically engineered host cells comprising expression systems. Accordingly, in a further aspect, the present invention relates to expression systems which comprise a polynucleotide or polynucleotides of the present invention, to host cells which are genetically engineered with such expression sytems and to the production of polypeptides of the invention by recombinant techniques. Cell-free translation systems can also be employed to produce such proteins using RNAs derived from the DNA constructs of the present invention.

For recombinant production, host cells can be genetically engineered to incorporate expression systems or portions thereof for polynucleotides of the present invention. Introduction of polynucleotides into host cells can be effected by methods described in many standard laboratory manuals, such as Davis et al, Basic Methods in Molecular Biology (1986) and Sambrook *et al.*, Molecular Cloning: A Laboratory Manual, 2nd Ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. (1989). Preferred such methods include, for instance, calcium phosphate transfection, DEAE-dextran mediated transfection, transvection, microinjection, cationic lipid-mediated transfection, electroporation, transduction, scrape loading, ballistic introduction or infection.

Representative examples of appropriate hosts include bacterial cells, such as *streptococci, staphylococci, E. coli, Streptomyces* and *Bacillus subtilis* cells; fungal cells, such as yeast cells and *Aspergillus* cells; insect cells such as *Drosophila* S2 and *Spodoptera* Sf9 cells; animal cells such as CHO, COS, HeLa, C127, 3T3, BHK, HEK 293 and Bowes melanoma cells; and plant cells.

A great variety of expression systems can be used, for instance, chromosomal, episomal and virus-derived systems, e.g., vectors derived from bacterial plasmids, from bacteriophage, from transposons, from yeast episomes, from insertion elements, from yeast chromosomal elements, from viruses such as baculoviruses, papova viruses, such as SV40, vaccinia viruses, adenoviruses, fowl pox viruses, pseudorabies viruses and retroviruses, and vectors derived from combinations thereof, such as those derived from plasmid and bacteriophage genetic elements, such as cosmids and phagemids. The expression systems may contain control regions that regulate as well as engender expression. Generally, any system or vector which is able to maintain, propagate or express a polynucleotide to produce a polypeptide in a host may be used. The appropriate nucleotide sequence may be inserted into an expression system by any of a variety of well-known and routine techniques, such as, for example, those set forth in Sambrook *et al.*, Molecular Cloning, A Laboratory Manual (supra). Appropriate secretion signals may be incorporated into the desired polypeptide to allow secretion of the translated protein into the lumen of the endoplasmic reticulum, the periplasmic space or the extracellular environment. These signals may be endogenous to the polypeptide or they may be heterologous signals.

If a polypeptide of the present invention is to be expressed for use in screening assays, it is generally preferred that the polypeptide be produced at the surface of the cell. In this event, the cells may be harvested prior to use in the screening assay. If the polypeptide is secreted into the medium, the medium can be recovered in order to recover and purify the polypeptide. If produced intracellularly, the cells must first be lysed before the polypeptide is recovered.

Polypeptides of the present invention can be recovered and purified from recombinant cell cultures by well-known methods including ammonium sulfate or ethanol precipitation, acid extraction, anion or cation exchange chromatography, phosphocellulose chromatography, hydrophobic interaction chromatography, affinity chromatography, hydroxylapatite chromatography and lectin chromatography. Most preferably, high performance liquid chromatography is employed for purification. Well known techniques for refolding proteins may be employed to regenerate active conformation when the polypeptide is denatured during isolation and or purification.

Another important aspect of the invention relates to a method for inducing , re-inforcing or modulating an immunological response in a mammal which comprises inoculating the mammal with a fragment or the entire polypeptide or polynucleotide of the present invention,

adequate to produce antibody and/or T cell immune response for prophylaxis or for therapeutic treatment of the diseases hereinbefore mentioned, amongst others. Yet another aspect of the invention relates to a method of inducing, re-inforcing or modulating immunological response in a mammal which comprises, delivering a polypeptide of the present invention *via* a vector or cell directing expression of the polynucleotide and coding for the polypeptide *in vivo* in order to induce such an immunological response to produce immune responses for prophylaxis or treatment of said mammal from diseases.

A further aspect of the invention relates to an immunological/vaccine formulation (composition) which, when introduced into a mammalian host, induces, re-inforces or modulates an immunological response in that mammal to a polypeptide of the present invention wherein the composition comprises a fragment or the entire polypeptide or polynucleotide of the present invention. The vaccine formulation may further comprise a suitable carrier. Since a polypeptide may be broken down in the stomach, it is preferably administered parenterally (for instance, subcutaneous, intramuscular, intravenous, or intradermal injection). Formulations suitable for parenteral administration include aqueous and non-aqueous sterile injection solutions which may contain anti-oxidants, buffers, bacteriostats and solutes which render the formulation isotonic with the blood of the recipient; and aqueous and non-aqueous sterile suspensions which may include suspending agents or thickening agents. The formulations may be presented in unit-dose or multi-dose containers, for example, sealed ampoules and vials and may be stored in a freeze-dried condition requiring only the addition of the sterile liquid carrier immediately prior to use. The vaccine formulation may also include adjuvant systems for enhancing the immunogenicity of the formulation, such as oil-in water systems, immunomodulators and other systems known in the art. The dosage will depend on the specific activity of the vaccine and can be readily determined by routine experimentation.

A further aspect of the invention relates to the *in vitro* induction of immune responses to a fragment or the entire polypeptide or polynucleotide of the present invention or a molecule comprising the polypeptide or polynucleotide of the present invention, using cells from the immune system of a mammal, and reinfusing these activated immune cells of the mammal for the treatment of disease. Activation of the cells from the immune system is achieved by in vitro incubation with the entire polypeptide or polynucleotide of the present invention or a molecule comprising the polypeptide or polynucleotide of the present invention in the presence or absence of various immunomodulator molecules.

A further aspect of the invention relates to the immunization of a mammal by administration of antigen presenting cells modified by *in vitro* loading with part or the entire

polypeptide of the present invention or a molecule comprising the polypeptide of the present invention and administered *in vivo* in an immunogenic way. Alternatively, antigen presenting cells can be transfected *in vitro* with a vector containing a fragment or the entire polynucleotide of the present invention or a molecule comprising the polynucleotide of the present invention, such as to express the corresponding polypeptide, and administered *in vivo* in an immunogenic way.

This invention also relates to the use of polynucleotides of the present invention as diagnostic reagents. Detection of a mutated form of the gene characterised by the polynucleotide of SEQ ID NO:1 which is associated with a dysfunction will provide a diagnostic tool that can add to, or define, a diagnosis of a disease, or susceptibility to a disease, which results from under-expression, over-expression or altered expression of the gene. Individuals carrying mutations in the gene may be detected at the DNA level by a variety of techniques.

Nucleic acids for diagnosis may be obtained from a subject's cells, such as from blood, urine, saliva, tissue biopsy or autopsy material. The genomic DNA may be used directly for detection or may be amplified enzymatically by using PCR or other amplification techniques prior to analysis. RNA or cDNA may also be used in similar fashion. Deletions and insertions can be detected by a change in size of the amplified product in comparison to the normal genotype. Point mutations can be identified by hybridizing amplified DNA to labeledCASB7439 nucleotide sequences. Perfectly matched sequences can be distinguished from mismatched duplexes by RNase digestion or by differences in melting temperatures. DNA sequence differences may also be detected by alterations in electrophoretic mobility of DNA fragments in gels, with or without denaturing agents, or by direct DNA sequencing (ee, e.g., Myers *et al.*, Science (1985) 230:1242). Sequence changes at specific locations may also be revealed by nuclease protection assays, such as RNase and S1 protection or the chemical cleavage method (see Cotton *et al.*, Proc Natl Acad Sci USA (1985) 85: 4397-4401). In another embodiment, an array of oligonucleotides probes comprising CASB7439 nucleotide sequence or fragments thereof can be constructed to conduct efficient screening of e.g., genetic mutations. Array technology methods are well known and have general applicability and can be used to address a variety of questions in molecular genetics including gene expression, genetic linkage, and genetic variability (see for example: M.Chee et al., Science, Vol 274, pp 610-613 (1996)).

The diagnostic assays offer a process for diagnosing or determining a susceptibilityto the Diseases through detection of mutation in the CASB7439 gene by the methods described. In addition, such diseases may be diagnosed by methods comprising determining from a sample derived from a subject an abnormally decreased or increased level of polypeptide or mRNA.

Decreased or increased expression can be measured at the RNA level using any of the methods well known in the art for the quantitation of polynucleotides, such as, for example, nucleic acid amplification, for instance PCR, RT-PCR, RNase protection, Northern blotting and other hybridization methods. Assay techniques that can be used to determine levels of a protein, such as a polypeptide of the present invention, in a sample derived from a host are well-known to those of skill in the art. Such assay methods include radioimmunoassays, competitive-binding assays, Western Blot analysis and ELISA assays.

Thus in another aspect, the present invention relates to a diagnostic kit which comprises:
(a) a polynucleotide of the present invention, preferably the nucleotide sequence of SEQ ID NO: 1, or a fragment thereof;
(b) a nucleotide sequence complementary to that of (a);
(c) a polypeptide of the present invention; or
(d) an antibody to a polypeptide of the present invention.

It will be appreciated that in any such kit, (a), (b), (c) or (d) may comprise a substantial component. Such a kit will be of use in diagnosing a disease or suspectability to a disease, particularly, amongst others.

The nucleotide sequences of the present invention are also valuable for chromosome identification. The sequence is specifically targeted to, and can hybridize with, a particular location on an individual human chromosome. The mapping of relevant sequences to chromosomes according to the present invention is an important first step in correlating those sequences with gene associated disease. Once a sequence has been mapped to a precise chromosomal location, the physical position of the sequence on the chromosome can be correlated with genetic map data. Such data are found in, for example, V. McKusick, Mendelian Inheritance in Man (available on-line through Johns Hopkins University Welch Medical Library). The relationship between genes and diseases that have been mapped to the same chromosomal region are then identified through linkage analysis (coinheritance of physically adjacent genes).

The differences in the cDNA or genomic sequence between affected and unaffected individuals can also be determined. If a mutation is observed in some or all of the affected individuals but not in any normal individuals, then the mutation is likely to be the causative agent of the disease.

The polypeptides of the invention or their fragments or analogs thereof, or cells expressing them, can also be used as immunogens to produce antibodies immunospecific for polypeptides of

the present invention. The term "immunospecific" means that the antibodies have substantially greater affinity for the polypeptides of the invention than their affinity for other related polypeptides in the prior art.

Antibodies generated against polypeptides of the present invention may be obtained by administering the polypeptides or epitope-bearing fragments, analogs or cells to an animal, preferably a non-human animal, using routine protocols. For preparation of monoclonal antibodies, any technique which provides antibodies produced by continuous cell line cultures can be used. Examples include the hybridoma technique (Kohler, G. and Milstein, C., Nature (1975) 256:495-497), the trioma technique, the human B-cell hybridoma technique (Kozbor *et al.*, Immunology Today (1983) 4:72) and the EBV-hybridoma technique (Cole *et al.*, Monoclonal Antibodies and Cancer Therapy, 77-96, Alan R. Liss, Inc., 1985).

Techniques for the production of single chain antibodies, such as those described in U.S. Patent No. 4,946,778, can also be adapted to produce single chain antibodies to polypeptides of this invention. Also, transgenic mice, or other organisms, including other mammals, may be used to express humanized antibodies.

The above-described antibodies may be employed to isolate or to identify clones expressing the polypeptide or to purify the polypeptides by affinity chromatography.

Antibodies against polypeptides of the present invention may also be employed to treat the Diseases, amongst others.

In a further aspect, the present invention relates to genetically engineered soluble fusion proteins comprising a polypeptide of the present invention, or a fragment thereof, and various portions of the constant regions of heavy or light chains of immunoglobulins of various subclasses (IgG, IgM, IgA, IgE). Preferred as an immunoglobulin is the constant part of the heavy chain of human IgG, particularly IgG1, where fusion takes place at the hinge region. In a particular embodiment, the Fc part can be removed simply by incorporation of a cleavage sequence which can be cleaved with blood clotting factor Xa. Furthermore, this invention relates to processes for the preparation of these fusion proteins by genetic engineering, and to the use thereof for drug screening, diagnosis and therapy. A further aspect of the invention also relates to polynucleotides encoding such fusion proteins. Examples of fusion protein technology can be found in International Patent Application Nos. WO94/29458 and WO94/22914.

Another aspect of the invention relates to a method for inducing an immunological response in a mammal which comprises inoculating the mammal with a polypeptide of the present invention, adequate to produce antibody and/or T cell immune response to protect said animal from the Diseases hereinbefore mentioned, amongst others. Yet another aspect of the

invention relates to a method of inducing immunological response in a mammal which comprises, delivering a polypeptide of the present invention *via* a vector directing expression of the polynucleotide and coding for the polypeptide *in vivo* in order to induce such an immunological response to produce antibody to protect said animal from diseases.

Polypeptides of the present invention are responsible for many biological functions, including many disease states, in particular the Diseases hereinbefore mentioned. It is therefore desirous to devise screening methods to identify compounds which stimulate or which inhibit the function of the polypeptide. Accordingly, in a further aspect, the present invention provides for a method of screening compounds to identify those which stimulate or which inhibit the function of the polypeptide. In general, agonists or antagonists may be employed for therapeutic and prophylactic purposes for such Diseases as hereinbefore mentioned. Compounds may be identified from a variety of sources, for example, cells, cell-free preparations, chemical libraries, and natural product mixtures. Such agonists, antagonists or inhibitors so-identified may be natural or modified substrates, ligands, receptors, enzymes, etc., as the case may be, of the polypeptide; or may be structural or functional mimetics thereof (see Coligan*et al.*, Current Protocols in Immunology 1(2):Chapter 5 (1991)).

The screening method may simply measure the binding of a candidate compound to the polypeptide, or to cells or membranes bearing the polypeptide, or a fusion protein thereof by means of a label directly or indirectly associated with the candidate compound. Alternatively, the screening method may involve competition with a labeled competitor. Further, these screening methods may test whether the candidate compound results in a signal generated by activation or inhibition of the polypeptide, using detection systems appropriate to the cells bearing the polypeptide. Inhibitors of activation are generally assayed in the presence of a known agonist and the effect on activation by the agonist by the presence of the candidate compound is observed. Constitutively active polpypeptides may be employed in screening methods for inverse agonists or inhibitors, in the absence of an agonist or inhibitor, by testing whether the candidate compound results in inhibition of activation of the polypeptide. Further, the screening methods may simply comprise the steps of mixing a candidate compound with a solution containing a polypeptide of the present invention, to form a mixture, measuring CASB7439 activity in the mixture, and comparing the CASB7439 activity of the mixture to a standard. Fusion proteins, such as those made from Fc portion and CASB7439 polypeptide, as hereinbefore described, can also be used for high-throughput screening assays to identify antagonists for the polypeptide of the present invention (see D. Bennett *et al.*, J Mol Recognition, 8:52-58 (1995); and K. Johanson *et al.*, J Biol Chem, 270(16):9459-9471 (1995)).

The polynucleotides, polypeptides and antibodies to the polypeptide of the present invention may also be used to configure screening methods for detecting the effect of added compounds on the production of mRNA and polypeptide in cells. For example, an ELISA assay may be constructed for measuring secreted or cell associated levels of polypeptide using monoclonal and polyclonal antibodies by standard methods known in the art. This can be used to discover agents which may inhibit or enhance the production of polypeptide (also called antagonist or agonist, respectively) from suitably manipulated cells or tissues.

The polypeptide may be used to identify membrane bound or soluble receptors, if any, through standard receptor binding techniques known in the art. These include, but are not limited to, ligand binding and crosslinking assays in which the polypeptide is labeled with a radioactive isotope (for instance, ^{125}I), chemically modified (for instance, biotinylated), or fused to a peptide sequence suitable for detection or purification, and incubated with a source of the putative receptor (cells, cell membranes, cell supernatants, tissue extracts, bodily fluids). Other methods include biophysical techniques such as surface plasmon resonance and spectroscopy. These screening methods may also be used to identify agonists and antagonists of the polypeptide which compete with the binding of the polypeptide to its receptors, if any. Standard methods for conducting such assays are well understood in the art.

Examples of potential polypeptide antagonists include antibodies or, in some cases, oligonucleotides or proteins which are closely related to the ligands, substrates, receptors, enzymes, etc., as the case may be, of the polypeptide, e.g., a fragment of the ligands, substrates, receptors, enzymes, etc.; or small molecules which bind to the polypetide of the present invention but do not elicit a response, so that the activity of the polypeptide is prevented.

Thus, in another aspect, the present invention relates to a screening kit for identifying agonists, antagonists, ligands, receptors, substrates, enzymes, etc. for polypeptides of the present invention; or compounds which decrease or enhance the production of such polypeptides, which comprises:

(a) a polypeptide of the present invention;

(b) a recombinant cell expressing a polypeptide of the present invention;

(c) a cell membrane expressing a polypeptide of the present invention; or

(d) antibody to a polypeptide of the present invention.;

It will be appreciated that in any such kit, (a), (b), (c) or (d) may comprise a substantial component.

It will be readily appreciated by the skilled artisan that a polypeptide of the present invention may also be used in a method for the structure-based design of an agonist, antagonist or inhibitor of the polypeptide, by:

(a) determining in the first instance the three-dimensional structure of the polypeptide;

(b) deducing the three-dimensional structure for the likely reactive or binding site(s) of an agonist, antagonist or inhibitor;

(c) synthesing candidate compounds that are predicted to bind to or react with the deduced binding or reactive site; and

(d) testing whether the candidate compounds are indeed agonists, antagonists or inhibitors.

It will be further appreciated that this will normally be an interative process.

In a further aspect, the present invention provides methods of treating abnormal conditions such as, for instance, cancer and autoimmune diseases, related to either an excess of, or an under-expression of, CASB7439 polypeptide activity.

If the activity of the polypeptide is in excess, several approaches are available. One approach comprises administering to a subject in need thereof an inhibitor compound (antagonist) as hereinabove described, optionally in combination with a pharmaceutically acceptable carrier, in an amount effective to inhibit the function of the polypeptide, such as, for example, by blocking the binding of ligands, substrates, receptors, enzymes, etc., or by inhibiting a second signal, and thereby alleviating the abnormal condition. In another approach, soluble forms of the polypeptides still capable of binding the ligand, substrate, enzymes, receptors, etc. in competition with endogenous polypeptide may be administered. Typical examples of such competitors include fragments of the CASB7439 polypeptide.

In still another approach, expression of the gene encoding endogenous CASB7439 polypeptide can be inhibited using expression blocking techniques. Known such techniques involve the use of antisense sequences, either internally generated or separately administered (see, for example, O'Connor, J Neurochem (1991) 56:560 in Oligodeoxynucleotides as Antisense Inhibitors of Gene Expression, CRC Press, Boca Raton, FL (1988)). Alternatively, oligonucleotides which form triple helices with the gene can be supplied (see, for example, Lee *et al.*, Nucleic Acids Res (1979) 6:3073; Cooney *et al.*, Science (1988) 241:456; Dervan *et al.*, Science (1991) 251:1360). These oligomers can be administered *per se* or the relevant oligomers can be expressed *in vivo*.

For treating abnormal conditions related to an under-expression of CASB7439 and its activity, several approaches are also available. One approach comprises administering to a subject a therapeutically effective amount of a compound which activates a polypeptide of the present

invention, i.e., an agonist as described above, in combination with a pharmaceutically acceptable carrier, to thereby alleviate the abnormal condition. Alternatively, gene therapy may be employed to effect the endogenous production of CASB7439 by the relevant cells in the subject. For example, a polynucleotide of the invention may be engineered for expression in a replication defective retroviral vector, as discussed above. The retroviral expression construct may then be isolated and introduced into a packaging cell transduced with a retroviral plasmid vector containing RNA encoding a polypeptide of the present invention such that the packaging cell now produces infectious viral particles containing the gene of interest. These producer cells may be administered to a subject for engineering cells *in vivo* and expression of the polypeptide *in vivo*. For an overview of gene therapy, see Chapter 20, Gene Therapy and other Molecular Genetic-based Therapeutic Approaches, (and references cited therein) in Human Molecular Genetics, T Strachan and A P Read, BIOS Scientific Publishers Ltd (1996). Another approach is to administer a therapeutic amount of a polypeptide of the present invention in combination with a suitable pharmaceutical carrier.

In a further aspect, the present invention provides for pharmaceutical compositions comprising a therapeutically effective amount of a polypeptide, such as the soluble form of a polypeptide of the present invention, agonist/antagonist peptide or small molecule compound, in combination with a pharmaceutically acceptable carrier or excipient. Such carriers include, but are not limited to, saline, buffered saline, dextrose, water, glycerol, ethanol, and combinations thereof. The invention further relates to pharmaceutical packs and kits comprising one or more containers filled with one or more of the ingredients of the aforementioned compositions of the invention. Polypeptides and other compounds of the present invention may be employed alone or in conjunction with other compounds, such as therapeutic compounds.

The composition will be adapted to the route of administration, for instance by a systemic or an oral route. Preferred forms of systemic administration include injection, typically by intravenous injection. Other injection routes, such as subcutaneous, intramuscular, or intraperitoneal, can be used. Alternative means for systemic administration include transmucosal and transdermal administration using penetrants such as bile salts or fusidic acids or other detergents. In addition, if a polypeptide or other compounds of the present invention can be formulated in an enteric or an encapsulated formulation, oral administration may also be possible. Administration of these compounds may also be topical and/or localized, in the form of salves, pastes, gels, and the like.

The dosage range required depends on the choice of peptide or other compounds of the present invention, the route of administration, the nature of the formulation, the nature of the subject's condition, and the judgment of the attending practitioner. Suitable dosages, however, are in the range of 0.1-100 µg/kg of subject. Wide variations in the needed dosage, however, are to be

expected in view of the variety of compounds available and the differing efficiencies of various routes of administration. For example, oral administration would be expected to require higher dosages than administration by intravenous injection. Variations in these dosage levels can be adjusted using standard empirical routines for optimization, as is well understood in the art.

Polypeptides used in treatment can also be generated endogenously in the subject, in treatment modalities often referred to as "gene therapy" as described above. Thus, for example, cells from a subject may be engineered with a polynucleotide, such as a DNA or RNA, to encode a polypeptide *ex vivo*, and for example, by the use of a retroviral plasmid vector. The cells are then introduced into the subject.

Polynucleotide and polypeptide sequences form a valuable information resource with which to identify further sequences of similar homology. This is most easily facilitated by storing the sequence in a computer readable medium and then using the stored data to search a sequence database using well known searching tools, such as GCC. Accordingly, in a further aspect, the present invention provides for a computer readable medium having stored thereon a polynucleotide comprising the sequence of SEQ ID NO:1 and/or a polypeptide sequence encoded thereby.

The following definitions are provided to facilitate understanding of certain terms used frequently hereinbefore.

"Antibodies" as used herein includes polyclonal and monoclonal antibodies, chimeric, single chain, and humanized antibodies, as well as Fab fragments, including the produots of an Fab or other immunoglobulin expression library.

"Isolated" means altered "by the hand of man" from the natural state. If an "isolated" composition or substance occurs in nature, it has been changed or removed from its original environment, or both. For example, a polynucleotide or a polypeptide naturally present in a living animal is not "isolated," but the same polynucleotide or polypeptide separated from the coexisting materials of its natural state is "isolated", as the term is employed herein.

"Polynucleotide" generally refers to any polyribonucleotide or polydeoxribonucleotide, which may be unmodified RNA or DNA or modified RNA or DNA. "Polynucleotides" include, without limitation, single- and double-stranded DNA, DNA that is a mixture of single- and double-stranded regions, single- and double-stranded RNA, and RNA that is mixture of single- and double-stranded regions, hybrid molecules comprising DNA and RNA that may be single-stranded or, more typically, double-stranded or a mixture of single- and double-stranded regions. In addition, "polynucleotide" refers to triple-stranded regions comprising RNA or DNA or both

RNA and DNA. The term "polynucleotide" also includes DNAs or RNAs containing one or more modified bases and DNAs or RNAs with backbones modified for stability or for other reasons. "Modified" bases include, for example, tritylated bases and unusual bases such as inosine. A variety of modifications may be made to DNA and RNA; thus, "polynucleotide" embraces chemically, enzymatically or metabolically modified forms of polynucleotides as typically found in nature, as well as the chemical forms of DNA and RNA characteristic of viruses and cells. "Polynucleotide" also embraces relatively short polynucleotides, often referred to as oligonucleotides.

"Polypeptide" refers to any peptide or protein comprising two or more amino acids joined to each other by peptide bonds or modified peptide bonds, i.e., peptide isosteres. "Polypeptide" refers to both short chains, commonly referred to as peptides, oligopeptides or oligomers, and to longer chains, generally referred to as proteins. Polypeptides may contain amino acids other than the 20 gene-encoded amino acids. "Polypeptides" include amino acid sequences modified either by natural processes, such as post-translational processing, or by chemical modification techniques which are well known in the art. Such modifications are well described in basic texts and in more detailed monographs, as well as in a voluminous research literature. Modifications may occur anywhere in a polypeptide, including the peptide backbone, the amino acid side-chains and the amino or carboxyl termini. It will be appreciated that the same type of modification may be present to the same or varying degrees at several sites in a given polypeptide. Also, a given polypeptide may contain many types of modifications. Polypeptides may be branched as a result of ubiquitination, and they may be cyclic, with or without branching. Cyclic, branched and branched cyclic polypeptides may result from post-translation natural processes or may be made by synthetic methods. Modifications include acetylation, acylation, ADP-ribosylation, amidation, covalent attachment of flavin, covalent attachment of a heme moiety, covalent attachment of a nucleotide or nucleotide derivative, covalent attachment of a lipid or lipid derivative, covalent attachment of phosphotidylinositol, cross-linking, cyclization, disulfide bond formation, demethylation, formation of covalent cross-links, formation of cystine, formation of pyroglutamate, formylation, gamma-carboxylation, glycosylation, GPI anchor formation, hydroxylation, iodination, methylation, myristoylation, oxidation, proteolytic processing, phosphorylation, prenylation, racemization, selenoylation, sulfation, transfer-RNA mediated addition of amino acids to proteins such as arginylation, and ubiquitination (see, for instance, Proteins - Structure and Molecular Properties, 2nd Ed., T. E. Creighton, W. H. Freeman and Company, New York, 1993; Wold, F., Post-translational Protein Modifications: Perspectives and Prospects, pgs. 1-12 in Post-translational Covalent Modification of Proteins, B. C. Johnson, Ed.,

Academic Press, New York, 1983; Seifter *et al.*, "Analysis for protein modifications and nonprotein cofactors", Meth Enzymol (1990) 182:626-646 and Rattan *et al.*, "Protein Synthesis: Post-translational Modifications and Aging", Ann NY Acad Sci (1992) 663:48-62).

"Variant" refers to a polynucleotide or polypeptide that differs from a reference polynucleotide or polypeptide, but retains essential properties. A typical variant of a polynucleotide differs in nucleotide sequence from another, reference polynucleotide. Changes in the nucleotide sequence of the variant may or may not alter the amino acid sequence of a polypeptide encoded by the reference polynucleotide. Nucleotide changes may result in amino acid substitutions, additions, deletions, fusions and truncations in the polypeptide encoded by the reference sequence, as discussed below. A typical variant of a polypeptide differs in amino acid sequence from another, reference polypeptide. Generally, differences are limited so that the sequences of the reference polypeptide and the variant are closely similar overall and, in many regions, identical. A variant and reference polypeptide may differ in amino acid sequence by one or more substitutions, additions, deletions in any combination. A substituted or inserted amino acid residue may or may not be one encoded by the genetic code. A variant of a polynucleotide or polypeptide may be a naturally occurring such as an allelic variant, or it may be a variant that is not known to occur naturally. Non-naturally occurring variants of polynucleotides and polypeptides may be made by mutagenesis techniques or by direct synthesis.

"Identity," as known in the art, is a relationship between two or more polypeptide sequences or two or more polynucleotide sequences, as determined by comparing the sequences. In the art, "identity" also means the degree of sequence relatedness between polypeptide or polynucleotide sequences, as the case may be, as determined by the match between strings of such sequences. "Identity" and "similarity" can be readily calculated by known methods, including but not limited to those described in (Computational Molecular Biology, Lesk, A.M., ed., Oxford University Press, New York, 1988; Biocomputing: Informatics and Genome Projects, Smith, D.W., ed., Academic Press, New York, 1993; Computer Analysis of Sequence Data, Part I, Griffin, A.M., and Griffin, H.G., eds., Humana Press, New Jersey, 1994; Sequence Analysis in Molecular Biology, von Heinje, G., Academic Press, 1987; and Sequence Analysis Primer, Gribskov, M. and Devereux, J., eds., M Stockton Press, New York, 1991; and Carillo, H., and Lipman, D., SIAM *J. Applied Math., 48:* 1073 (1988). Preferred methods to determine identity are designed to give the largest match between the sequences tested. Methods to determine identity and similarity are codified in publicly available computer programs. Preferred computer program methods to determine identity and similarity between two sequences include, but are not limited to, the GCG program package (Devereux, J., et al., Nucleic Acids Research *12(1):* 387 (1984)), BLASTP, BLASTN,

and FASTA (Atschul, S.F. et al., J. Molec. Biol. *215*: 403-410 (1990). The BLAST X program is publicly available from NCBI and other sources (BLAST Manual, Altschul, S., *et al.*, NCBI NLM NIH Bethesda, MD 20894; Altschul, S., *et al.*, J. Mol. Biol. *215*: 403-410 (1990). The well known Smith Waterman algorithm may also be used to determine identity.

Preferred parameters for polypeptide sequence comparison include the following:

1) Algorithm: Needleman and Wunsch, J. Mol Biol. 48: 443-453 (1970)

Comparison matrix: BLOSSUM62 from Hentikoff and Hentikoff, Proc. Natl. Acad. Sci. USA. 89:10915-10919 (1992)

Gap Penalty: 12

Gap Length Penalty: 4

A program useful with these parameters is publicly available as the "gap" program from Genetics Computer Group, Madison WI. The aforementioned parameters are the default parameters for polypeptide comparisons (along with no penalty for end gaps).

Preferred parameters for polynucleotide comparison include the following:

1) Algorithm: Needleman and Wunsch, J. Mol Biol. 48: 443-453 (1970)

Comparison matrix: matches = +10, mismatch = 0

Gap Penalty: 50

Gap Length Penalty: 3

A program useful with these parameters is publicly available as the "gap" program from Genetics Computer Group, Madison WI. The aforementioned parameters are the default parameters for polynucleotide comparisons.

By way of example, a polynucleotide sequence of the present invention may be identical to the reference sequence of SEQ ID NO:1, that is be 100% identical, or it may include up to a certain integer number of nucleotide alterations as compared to the reference sequence. Such alterations are selected from the group consisting of at least one nucleotide deletion, substitution, including transition and transversion, or insertion, and wherein said alterations may occur at the 5' or 3' terminal positions of the reference nucleotide sequence or anywhere between those terminal positions, interspersed either individually among the nucleotides in the reference sequence or in one or more contiguous groups within the reference sequence. The number of nucleotide alterations is determined by multiplying the total number of nucleotides in SEQ ID NO:1 by the numerical percent of the respective percent identity(divided by 100) and subtracting that product from said total number of nucleotides in SEQ ID NO:1, or:

$$n_n \leq x_n - (x_n * y),$$

wherein n_n is the number of nucleotide alterations, x_n is the total number of nucleotides in SEQ ID NO:1,

"Fusion protein" refers to a protein encoded by two, often unrelated, fused genes or fragments thereof. In one example, EP-A-0 464 discloses fusion proteins comprising various portions of constant region of immunoglobulin molecules together with another human protein or part thereof. In many cases, employing an immunoglobulin Fc region as a part of a fusion protein is advantageous for use in therapy and diagnosis resulting in, for example, improved pharmacokinetic properties [see, e.g., EP-A 0232 262]. On the other hand, for some uses it would be desirable to be able to delete the Fc part after the fusion protein has been expressed, detected and purified.

All publications, including but not limited to patents and patent applications, cited in this specification are herein incorporated by reference as if each individual publication were specifically and individually indicated to be incorporated by reference herein as though fully set forth.

Examples

Subtractive cDNA cloning of colon tumour-associated antigen (TAA) candidates.

Subtractive cDNA libraries are produced using standard technologies. Briefly, total RNA is extracted from frozen (-70°C) tumour and matched normal colon samples using the TriPure reagent and protocol (Boehringer). Target RNA is prepared by pooling total RNA from three tumour samples (30 µg each). Driver RNA is prepared by pooling total RNA from three matched normal colon samples (10 µg each) and total RNA from seven normal tissues other than colon(brain, heart, kidney, liver, bladder, skin, spleen; 10 µg each). Total RNA from non-colon normal tissues is purchased from InVitrogen.

Messenger RNA is purified from total RNA using oligo-dT magnetic bead technology (Dynal) and quantified by spectrofluorimetry (BioRad).

Target and driver mRNA are reverse transcribed into cDNA using one of two strategies: 1) Target sequences for PCR oligonucleotides are introduced onto the ends of the newly synthesised cDNA during reverse transcription using the template switching capability of reverse transcriptase (ClonTech SMART PCR cDNA synthesis kit). 2) Alternatively, the target and driver mRNA are reverse transcribed into cDNA using an oligo-dT primer and converted to double-strand cDNA; the

cDNA is cleaved with RsaI and linkers for PCR amplification are ligated onto the extremities of the cDNA fragments.

In both cases, target and driver cDNA are amplified by long range PCR (ClonTech SMART PCR Synthesis Kit and Advantage PCR Polymerase Mix) and used as starting material for subtractive cloning. For amplification, cycling conditions and optimisation of the number of PCR cycles are as described in the Advantage PCR protocol.

Two subtractive cloning strategies are used: ClonTech PCR SELECT (see ClonTech kit protocol and N. Gurskaya *et al.* 1996. Analytical Biochemistry: 240, 90) and cRDA(M. Hubank and D. Schatz. 1994. Nucleic Acids Research: 22, 5640) . When the PCR SELECT protocol is used, the primary PCR SELECT subtraction products are submitted to a supplementary round of cRDA subtraction. When the cRDA protocol is used, two consecutive cycles of cRDA subtraction are performed. In each case the products of both cycles of subtraction are cloned into pCR-TOPO (InVitrogen) and transformed into E. coli to produce a subtracted cDNA plasmid library.

An alternative strategy is also followed: subtraction of normal colon sequences and sequences from non-colon normal tissues are subtracted in separate hybridizations. In this case, target and driver RNA are assembled for the first subtraction as above with the exception that non-colon RNA is left out of the driver pool and amounts of normal colon are increased to 10 µg. Preparation of target and driver cDNA and subtractive hybridization are performed as described above. A second subtraction is then performed on the products of the first subtraction, but the driver is now composed of a pool of normal colon and normal non-colon mRNA from the seven normal tissues.

Differential Screening of cDNA arrays.

Identification of tumour-associated genes in the subtracted cDNA library is accomplished by differential screening.

Total bacterial DNA is extracted from 100 µl over-night cultures. Bacteria are lysed with guanidium isothiocyantate and the bacterial DNA is affinity purified using magnetic glass (Boehringer). Plasmid inserts are recovered from the bacterial DNA by Advantage PCR amplification (Clontech). The PCR products are dotted onto two nylon membranes to produce high density cDNA arrays using the Biomek 96 HDRT tool (Beckman). The spotted cDNA is covalently linked to the membrane by UV irradiation. The first membrane is hybridised with a mixed cDNA probe prepared from the tumour of a single patient. The second membrane is hybridised with an equivalent amount of mixed cDNA probe prepared from normal colon of the same patient. The

probe cDNA is prepared by PCR amplification as described above and is labelled using the AlkPhos Direct System (Amersham). Hybridisation conditions and stringency washes are as described in the AlkPhos Direct kit. Hybridized probe is detected by chemiluminescence. Hybridisation intensities for each cDNA fragment on both blots are measured by film densitometry or direct measurement (BioRad Fluor-S Max). The ratio of the tumour to normal hybridisation intensities (T/N) is calculated for each gene to evaluate the degree of over-expression in the tumour. Genes which are significantly over-expressed in colon tumours are followed-up. Significance is arbitrarily defined as one standard deviation of the T/N frequency distribution. Differential screening experiments are repeated using RNA from multiple patient donors (>18) to estimate the frequency of over-expressing tumours in the patient population.

In addition, the DNA arrays are hybridised with mixed cDNA probes from normal tissues other than colon (see list above) to determine the level of expression of the candidate gene in these tissues.

Real-time RT-PCR analysis

Real-time RT-PCR (Applied Biosystems 7700) is used to obtain an independent evaluation of the transcript abundance of a candidate antigen in tumour and normal tissue mRNA. This methodology uses minute amounts of mRNA as input material and is capable of detecting very low levels of transcripts with exquisite specificity (U. Gibson. 1996. Genome Research: 6,996). Real-time experiments are run on mRNA from tumour and matched normal colon. Ct values are obtained for each sample; the difference between these two values (delta Ct) is a direct measure of the difference in transcript levels between the tumor and normal tissue. Samples from multiple patient donors (>18) are analysed to estimate delta Ct values for a particular transcript in the patient population.

Northern-Southern blot analysis

Limited amounts of mixed tumor and matched normal colon cDNA are amplified by Advantage PCR (see above). Messenger RNA from multiple normal tissues is also amplified using the same procedure. The amplified cDNA (1 μg) is electrophoresed on a 1.2% agarose gel and transferred onto a nylon membrane. The membrane is hybridised (AlkPhos Direct System) with a probe prepared using a fragment of the candidate TAA cDNA. Northern-Southern analysis provides information on transcript size, presence of splice variants and transcript abundance in tumour and normal tissues. Northern-Southern Blot analysis is used in place of the more familiar Northern Blot analysis as most tissue samples are too small to yield enough RNA for the latter procedure.

SEQUENCE INFORMATION

SEQ ID NO:1

GGTAAACAGAACTGATTTATTTATAAAGTTCACGCTCCCTTGAAGAGGTGTGCCCCACAC
AGGCTTCTCCCTAGCAGAGCAGCAGTGCCCACAAACCCACCCCAGGGTGGGCTGTCACG
GGGGCCTCACGCCAGGGACCCCGCCCCTCAGGACTGCTCGTGTCCAGATCTTGGCCAGC
ATGGAAAACTCCAGATAGTGGGGGCAGGGGTCCAGGTCATCTTTATTACGCCC
CAGGTCAAGGGTTCTTTGTACAAAAATAGGTCTCCGTTTGCCAGCAGTGTCCCTCCAGCA
GCTCAAGTTAATGTGTAGAAAATGGATTCTCTGTGCCCTTAGAAAATCCTCTCCCCTCC
GGAAAAATCTCCAAGTGTTGGTGCCCCCCCCCCCACTGCAGTCGAGAAGCTGTGGGGAGG
GGCGGCGTCGGAGGAAGCCGCCAGCCCTTATGGGGCCAGCTCCAAGCCCGTTTCCACC
GCGGCATTGGTCAGGCTGGGCGGACGAACGAGGCGGCGTCGGCGGTGCGGGGGGTGGT
GGGTGGGTCCCCGGCTCGCTGGGGGCGGAGCAGCGGGCCGGTCCACCTGGCGGGCTCCCC

What is claimed is:

1. An isolated polynucleotide comprising a nucleotide sequence that has at least70% identity to a nucleotide sequence of SEQ ID NO:1 over its entire length; or a nucleotide sequence complementary to said isolated polynucleotide.

2. The polynucleotide of claim 1 which is DNA or RNA.

3. The polynucleotide of claim 1 which is the polynucleotide of SEQ ID NO: 1.

4. An isolated polynucleotide comprising a polynucleotide that hybridizes under stringent condition to a nucleotide sequence having SEQ ID NO: 1 or a fragment thereof.

5. A polypeptide comprising amino acid sequence encoded by the polynucleotide obtained in claim 4.

6. A vaccine comprising an effective amount of the polynucleotide of claim 1 and a pharmaceutically acceptable carrier.

7. A vaccine comprising an effective amount of the polypeptide of claim 5 and a pharmaceutically effective carrier.

8. Use of the vaccines as claimed in claims 6 and 7 in the immunoprophylaxis or therapy of disease.

Abstract

CASB7439 polypeptides and polynucleotides and methods for producing such polypeptides by recombinant techniques are disclosed. Also disclosed are methods for utilizing CASB7439 polypeptides and polynucleotides in diagnostics, prophylaxis and therapy, and diagnostic assays for such.

PCT/EP 01/01779

The Patent Office
Concept House
Cardiff Road
Newport
South Wales
NP10 8QQ

EPO - DG 1

09. 07. 2001

(83)

PRIORITY DOCUMENT

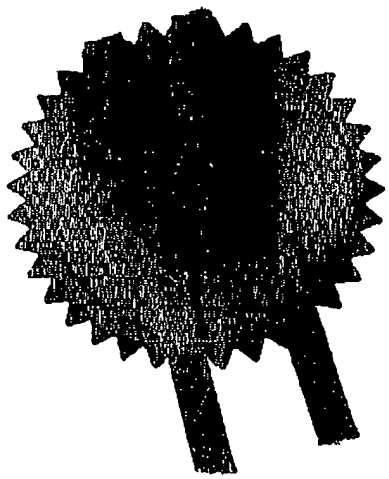
SUBMITTED OR TRANSMITTED IN COMPLIANCE WITH RULE 17.1(a) OR (b)

I, the undersigned, being an officer duly authorised in accordance with Section 74(1) and (4) of the Deregulation & Contracting Out Act 1994, to sign and issue certificates on behalf of the Comptroller-General, hereby certify that annexed hereto is a true copy of the documents as originally filed in connection with the patent application identified therein.

In accordance with the Patents (Companies Re-registration) Rules 1982, if a company named in this certificate and any accompanying documents has re-registered under the Companies Act 1980 with the same name as that with which it was registered immediately before re-registration save for the substitution as, or inclusion as, the last part of the name of the words "public limited company" or their equivalents in Welsh, references to the name of the company in this certificate and any accompanying documents shall be treated as references to the name with which it is so re-registered.

In accordance with the rules, the words "public limited company" may be replaced by p.l.c., plc, P.L.C. or PLC.

Re-registration under the Companies Act does not constitute a new legal entity but merely subjects the company to certain additional company law rules.



Signed A W Brewer.

Dated 1 June 2001

Patents Form 1/77

Patents Act 1977
(Rule 16)

The Patent Office

29AUG00 E564749-1 D02429
P01/7700 0.00-0021080.7

1/77

THE PATENT OFFICE
M
25 AUG 2000
LONDON

Request for grant of a patent

(See the notes on the back of this form. You can also get an explanatory leaflet from the Patent Office to help you fill in this form)

The Patent Office
Cardiff Road
Newport
Gwent NP9 1RH

1. Your reference — KLP/RH/BC45354

2. Patent application number
(The Patent Office will fill in his part) — 25 AUG 2000 — **0021080.7**

3. Full name, address and postcode of the or of each applicant *(underline all surnames)*

SmithKline Beecham Biologicals s.a.
Rue de l'Institut 89, B-1330 Rixensart, , Belgium

Patents ADP number *(if you know it)*

05781117001

If the applicant is a corporate body, give the country/state of its incorporation

Belgian

4. Title of the invention — Novel Compounds

5. Name of your agent *(if you have one)*

CORPORATE INTELLECTUAL PROPERTY

"Address for service" in the United Kingdom to which all correspondence should be sent
(including the postcode)

SMITHKLINE BEECHAM PLC
TWO NEW HORIZONS COURT
BRENTFORD
MIDDLESEX TW8 9EP

Patents ADP number *(if you know it)*

05776299003

6. If you are declaring priority from one or more earlier patent applications, give the country and the date of filing of the or each of these earlier applications and *(if you know it)* the or each application number

Country	Priority application number *(if you know it)*	Date of filing *(day / month / year)*

7. If this application is divided or otherwise derived from an earlier UK application, give the number and the filing date of the earlier application

Number of earlier application	Date of filing *(day / month / year)*

8. Is a statement of inventorship and of right to grant of a patent required in support of this request? *(Answer yes if:*

a) any applicant named in part 3 is not an inventor, or
b) there is an inventor who is named as an applicant, or
c) any named applicant is a corporate body
See note (d)

Patents Form 1/77

9. Enter the number of sheets for any of the following items you are filing with this form. Do not count copies of the same document

Continuation sheets of this form	
Description	46
Claim(s)	6
Abstract	1
Drawings	2 + 2

10. If you are also filing any of the following, state how many against each item.

Priority Documents	
Translations of priority documents	
Statement of inventorship and right to grant of a patent *(Patents Form 7/77)*	
Request for preliminary examination and search *(Patents Form 9/77)*	
Request for substantive examination *(Patents Form 10/77)*	
Any other documents *(please specify)*	

11. We request the grant of a patent on the basis of this application

Signature K L Privett Date 25-Aug-00

K L Privett

12. Name and daytime telephone number of person to contact in the United Kingdom

K L Privett 020 8975 2585

Warning

After an application for a Patent has beeen filed, the Comptroller of the Patent Office will consider whether publication or communication of the invention should be prohibited or restricted under Section 22 of the Patents Act 1977. You will be informed if it is necessary to prohibit or restrict your invention in this way. Furthermore, if you live in the United Kingdom, Section 23 of the Patents Act 1977 stops you from applying for a patent abroad without first getting written permission unless an application has been filed at least six weeks beforehand in the United Kingdom for a patent for the same invention and either no direction prohibiting publication or communication has been given, or any such direction has been revoked.

Notes

a) If you need help to fill in this form or you have any questions, please contact the Patent Office on 0645 500505

b) Write your answers in capital letters using black ink or you may type them.

c) If there is not enough space for all relevant details on any part of this form, please continue on a separate sheet of paper and write "see continuation sheet" in the relevant part(s). Any continuation sheet should be attached to this form.

d) If you have answered 'Yes' Patents Form 7/77 will need to be filed.
f) For details of the fee and ways to pay please contact the Patent Office.

Patents Form 1/77

Novel Compounds

Field of the Invention

This invention relates to newly identified polypeptides and polynucleotides encoding such polypeptides, to their use in diagnostics, prophylaxis and therapy and in identifying compounds which may be agonists, antagonists and /or inhibitors which are potentially useful in therapy, and to production of such polypeptides and polynucleotides. In particular these are useful for the treatment and diagnosis of colon carcinoma.

Background of the Invention

Colon cancer is the second most frequently diagnosed malignancy in the United Sates as well as the second most common cause of cancer death. An estimated 95,600 new cases of colon cancer will be diagnosed in 1998, with an estimated 47,700 deaths. The five-year survival rate for patients with colorectal cancer detected in an early-localised stage is 92%, unfortunately, only 37% of colorectal cancer is diagnosed at this stage. The survival stage. The survival rate drops to 64% if the cancer is allowed to spread to adjacent organs or lymph nodes, and to7% in patients with distant metastases.
The prognosis of colon cancer is directly related to the degree of penetration of the tumour through the bowel wall and the presence or absence of nodal involvement, consequently, early detection and treatment are especially important. Currently, diagnosis is aided by the use of screening assays for fecal occult blood, sigmoidoscopy, colonoscopy and double contrast barium enemas. Treatment regimens are determined by the type and stage of the cancer, and include surgery, radiation therapy and/or chemotherapy. Recurrence following surgery (the most common form of therapy) is a major problem and is often the ultimate cause of death. In spite of considerable research into therapies for the disease, colon cancer remains difficult to diagnose and treat.

Summary of the Invention

The present invention relates to CASB7439, in particular CASB7439 polypeptides and CASB7439 polynucleotides, recombinant materials and methods for their production. In another aspect, the invention relates to methods for using such polypeptides and polynucleotides, including the treatment of cancer and autoimmune diseases, hereinafter referred to as "the Diseases", amongst others. In a further aspect, the invention relates to methods for identifying agonists and

antagonists/inhibitors using the materials provided by the invention, and treating conditions associated with CASB7439 imbalance with the identified compounds In a still further aspect, the invention relates to diagnostic assays for detecting diseases associated with inappropriateCASB7439 activity or levels.

Description of the Invention

In a first aspect, the present invention relates to CASB7439 polypeptides. Such peptides include isolated polypetides comprising an amino acid sequence which has at least 70% identity, preferably at least 80% identity, more preferably at least 90% identity, yet more preferably at least 95% identity, most preferably at least 97-99% identity, to that of SEQ ID NO:2 and NO:3 over the entire length of SEQ ID NO:2 and SEQ ID NO:3, respectively. Such polypeptides include those comprising the amino acid of SEQ ID NO:2 and SEQ ID NO:3.

Further peptides of the present invention include isolated polypeptides in which the amino acid sequence has at least 70% identity, preferably at least 80% identity, more preferably at least 90% identity, yet more preferably at least 95% identity, most preferably at least 97-99% identity, to the amino acid sequence of SEQ ID NO:2 and SEQ ID NO:3 over the entire length of SEQ ID NO:2 and SEQ ID NO:3, respectively. Such polypeptides include the polypeptide of SEQ ID NO:2 and SEQ ID NO:3.

Further peptides of the present invention include isolated polypeptides encoded by a polynucleotide comprising the sequence contained in SEQ ID NO:1.

Polypeptides and polynucleotides of the present invention are believed to be important immunogens for specific prophylactic or therapeutic immunization against tumors, because they are highly overexpressed in tumors and can thus be targeted by antigen-specific immune mechanisms leading to the destruction of the tumor cell. They can also be used to diagnose the occurrence of tumor cells. Furthermore, their inappropriate expression in certain circumstances can cause an induction of autoimmune, inappropriate immune responses, which could be corrected through appropriate vaccination using the same polypeptides or polynucleotides. In this respect the most important biological activities to our purpose arethe antigenic and immunogenic activities of the polypeptide of the present invention. A polypeptide of the present invention may also exhibit at least one other biological activity of CASB7439, which could qualify it as a target for therapeutic or prophylactic intervention different from that linked to the immune response.

The invention also provides an immunogenic fragment of a CASB7439 polypeptide, that is a contiguous portion of the CASB7439 polypeptide. That is to say, the fragment (if necessary when coupled to a carrier) is capable of raising an immune response which recognises the

CASB7439 polypeptide. Such an immunogenic fragment may include, for example, the CASB7439 polypeptide lacking an N-terminal leader sequence, a transmembrane domain or a C-terminal anchor domain. In a preferred aspect the immunogenic fragment of CASB7439 according to the invention comprises substantially all of the extracellular domain of a polypeptide.

A fragment is a polypeptide having an amino acid sequence that is entirely the same as part but not all of any amino acid sequence of any polypeptide of the invention. As with CASB7439 polypeptides, fragments may be "free-standing", or comprised within a larger polypeptide of which they form a part or region, most preferably as a single continuous region in a single larger polypeptide.

Preferred fragments include, for example, truncation polypeptides having a portion of an amino acid sequence encoded by a CASB7439 polynucleotide or of a variant thereof, such as a continuous series of residues that includes an amino- and/or carboxyl-terminal amino acid sequence. Degradation forms of the polypeptides of the invention produced by or in a host cell, are also preferred. Further preferred are fragments characterised by structural or functional attributes such as fragments that comprise alpha-helix and alpha-helix forming regions, beta-sheet-forming regions, turn and turn-forming regions, coil and coil-forming regions, hydrophilic regions, hydrophobic regions, alpha amphipathic regions, beta amphiphathic regions, flexible regions, surface-forming regions, substrate binding region, and high antigenic index regions.

Further preferred fragments include an isolated polypeptide comprising an amino acid sequence having at least 15, 20, 30, 40, 50 or 100 contiguous amino acids from a CASB7439 polypeptide, or an isolated polypeptide comprising an amino acid sequence having at least 15, 20, 30, 40, 50 or 100 contiguous amino acids truncated or deleted from a CASB7439 polypeptide.

The polypeptides or immunogenic fragments of the present invention may be in the form of the "mature" protein or may be a part of a larger protein such as a fusion protein. It is often advantageous to include an additional amino acid sequence which contains secretory or leader sequences, pro-sequences, sequences which aid in purification such as multiple histidine residues, or an additional sequence for stability during recombinant production. Furthermore, addition of exogenous polypeptide or lipid tail or polynucleotide sequences to increase the immunogenic potential of the final molecule is also considered.

The present invention also includes variants of the aforementioned polypeptides, that is polypeptides that vary from the referents by conservative amino acid substitutions, whereby a residue is substituted by another with like characteristics. Typical such substitutions are among Ala, Val, Leu and Ile; among Ser and Thr; among the acidic residues Asp and Glu; among Asn and Gln; and

among the basic residues Lys and Arg; or aromatic residues Phe and Tyr. Particularly preferred are variants in which several, 5-10, 1-5, 1-3, 1-2 or 1 amino acids are substituted, deleted, or added in any combination.

Polypeptides of the present invention can be prepared in any suitable manner. Such polypeptides include isolated naturally occurring polypeptides, recombinantly produced polypeptides, synthetically produced polypeptides, or polypeptides produced by a combination of these methods. Means for preparing such polypeptides are well understood in the art.

In a further aspect, the present invention relates to CASB7439 polynucleotides. Such polynucleotides include isolated polynucleotides comprising a nucleotide sequence encoding a polypeptide which has at least 70% identity, preferably at least 80% identity, more preferably at least 90% identity, yet more preferably at least 95% identity, to the amino acid sequence of SEQ ID NO:2 and SEQ ID NO:3, over the entire length of SEQ ID NO:2 and SEQ ID NO:3, respectively. In this regard, polypeptides which have at least 97% identity are highly preferred, whilst those with at least 98-99% identity are more highly preferred, and those with at least 99% identity are most highly preferred. Such polynucleotides include a polynucleotide comprising the nucleotide sequence contained in SEQ ID NO:1 encoding the polypeptides of SEQ ID NO:2 and SEQ ID NO:3.

Further polynucleotides of the present invention include isolated polynucleotides comprising a nucleotide sequence that has at least 70% identity, preferably at least 80% identity, more preferably at least 90% identity, yet more preferably at least 95% identity, to a nucleotide sequence encoding a polypeptide of SEQ ID NO:2 and SEQ ID NO:3, over the entire coding region. In this regard, polynucleotides which have at least 97% identity are highly preferred, whilst those with at least 98-99% identity are more highly preferred, and those with at least 99% identity are most highly preferred.

Further polynucleotides of the present invention include isolated polynucleotidescomprising a nucleotide sequence which has at least 70% identity, preferably at least 80% identity, more preferably at least 90% identity, yet more preferably at least 95% identity, to SEQ ID NO:1 over the entire length of SEQ ID NO:1. In this regard, polynucleotides which have at least 97% identity are highly preferred, whilst those with at least 98-99% identity are more highly preferred, and those with at least 99% identity are most highly preferred. Such polynucleotides include a polynucleotide comprising the polynucleotide of SEQ ID NO:1 as well as the polynucleotide of SEQ ID NO:1. Said polynucleotide can be inserted in a suitable plasmid or recombinant microrganism vector and used for immunization (see for example Wolff et. al., Science 247:1465-1468 (1990); Corr et. al., J. Exp. Med. 184:1555-1560 (1996); Doe et. al., Proc. Natl. Acad. Sci. 93:8578-8583 (1996)).

The invention also provides polynucleotides which are complementary to all the above described polynucleotides.

The nucleotide sequence of SEQ ID NO:1 shows homology with Homo sapiens achaete-scute complex (Drosophila) homolog-like 2 (ASCL2) mRNA (genbank accession NM_005170). The nucleotide sequence of SEQ ID NO:1 is a cDNA sequence and comprises a polypeptide encoding sequence (nucleotide 545 to 1123) encoding a polypeptide of 193 amino acids, the polypeptide of SEQ ID NO:2. The nucleotide sequence encoding the polypeptide of SEQ ID NO:2 may be identical to the polypeptide encoding sequence contained in SEQ ID NO:1 or it may be a sequence other than the one contained in SEQ ID NO:1, which, as a result of the redundancy (degeneracy) of the genetic code, also encodes the polypeptide of SEQ ID NO:2. The polypeptide of the SEQ ID NO:2 is structurally related to other proteins of the Achaete-Scute protein family, having homology and/or structural similarity with human Achaete-Scute homologue 2 (genbank accession AAB86993).

Human Achaete Scute Homologue 2 (HASH2) gene, officially designated human ASCL2 (Achaete Scute complex like 2) is a homologue of the Drosophila Achaete and Scute genes. Human ASCL2 is expressed in the extravillus trophoblasts of the developing placenta only, and maps on chromosome 11p15 close to IGF2 and H19.

The mouse achaete-scute homolog-2 gene (MASH2) encodes a transcription factor playing a role in the development of the trophoblast.

The Mash2 gene is paternally imprinted in the mouse, and the lack of human ASCL2 expression in non-malignant hydatidiform (androgenetic) moles indicates that human Ascl2 is also imprinted in man.

Ascl2 genes are members of the basic helix-loop-helix (BHLH) family of transcription factors. They activate transcription by binding to the E box (5'-CANNTG-3'). Dimerization with other BHLH proteins is required for efficient DNA binding. They are involved in the determination of the neuronal precursors in the peripheral nervous system and the central nervous system in drosophila melanogaster, and probably in mammals as well.

The nucleotide sequence of SEQ ID NO:1 also comprises another polypeptide encoding sequence (nucleotide 840 to 262) encoding a polypeptide of 262 amino acids, the polypeptide of SEQ ID NO:3. The nucleotide sequence encoding the polypeptide of SEQ ID NO:3 may be identical to the polypeptide encoding sequence contained in SEQ ID NO:1 or it may be a sequence other than the one contained in SEQ ID NO:1, which, as a result of the redundancy (degeneracy) of the genetic code, also encodes the polypeptide of SEQ ID NO:3. The polypeptide of the SEQ ID NO3 is

structurally related to other proteins of the splicing coactivator protein family, having homology and/or structural similarity with homo sapiens splicing coactivator subunit srm300(genbank accession AAF21439).

Preferred polypeptides and polynucleotides of the present invention are expected to have, *inter alia*, similar biological functions/properties to their homologous polypeptides and polynucleotides. Furthermore, preferred polypeptides and polynucleotides of the present invention have at least one CASB7439 activity.

The present invention also relates to partial or other polynucleotide and polypeptide sequences which were first identified prior to the determination of the corresponding full length sequences of SEQ ID NO:1, SEQ ID NO:2 and SEQ ID NO:3.

Accordingly, in a further aspect, the present invention provides for an isolated polynucleotide comprising:

(a) a nucleotide sequence which has at least 70% identity, preferably at least 80% identity, more preferably at least 90% identity, yet more preferably at least 95% identity, even more preferably at least 97-99% identity to SEQ ID NO:4 and SEQ ID NO:5 over the entire length of SEQ ID NO:4 and SEQ ID NO:5, respectively;

(b) a nucleotide sequence which has at least 70% identity, preferably at least 80% identity, more preferably at least 90% identity, yet more preferably at least 95% identity, even more preferably at least 97-99% identity, to SEQ ID NO:1 over the entire length of SEQ ID NO:4 and SEQ ID NO:5;

(c) the polynucleotides of SEQ ID NO:4 and SEQ ID NO:5; or

(d) a nucleotide sequence encoding a polypeptide which has at least 70% identity, preferably at least 80% identity, more preferably at least 90% identity, yet more preferably at least 95% identity, even more preferably at least 97-99% identity, to the amino acid sequence of SEQ ID NO:6 and SEQ ID NO:7, over the entire length of SEQ ID NO:6 and SEQ ID NO:7, respectively; as well as the polynucleotide of SEQ ID NO:4 and SEQ ID NO:5.

The present invention further provides for a polypeptide which:

(a) comprises an amino acid sequence which has at least 70% identity, preferably at least 80% identity, more preferably at least 90% identity, yet more preferably at least 95% identity, most preferably at least 97-99% identity, to that of SEQ ID NO:2 and SEQ ID NO:3 over the entire length of SEQ ID NO:6 and SEQ ID NO:7, respectively;

(b) has an amino acid sequence which is at least 70% identity, preferably at least 80% identity, more preferably at least 90% identity, yet more preferably at least 95% identity, most preferably

at least 97-99% identity, to the amino acid sequence of SEQ ID NO:2 and SEQ ID NO:3 over the entire length of SEQ ID NO:6 and SEQ ID NO:7, respectively;

(c) comprises the amino acid of SEQ ID NO:6 and SEQ ID NO:7; and

(d) is the polypeptide of SEQ ID NO:6 and SEQ ID NO:7;

as well as polypeptides encoded by a polynucleotide comprising the sequence contained in SEQ ID NO:4 and SEQ ID NO:5.

The nucleotide sequence of SEQ ID NO:4 and SEQ ID NO:5 and the peptide sequence encoded thereby are derived from EST (Expressed Sequence Tag) sequences. It is recognised by those skilled in the art that there will inevitably be some nucleotide sequence reading errors in EST sequences (see Adams, M.D. *et al*, Nature 377 (supp) 3, 1995). Accordingly, the nucleotide sequence of SEQ ID NO:4 and SEQ ID NO:5 and the peptide sequence encoded therefrom are therefore subjec to the same inherent limitations in sequence accuracy. Furthermore, the peptide sequence encoded by SEQ ID NO:4 and SEQ ID NO:5 comprises a region of identity or close homology and/or close structural similarity (for example a conservative amino acid difference) with the closest homologous or structurally similar protein.

Polynucleotides of the present invention may be obtained, using standard cloning and screening techniques, from a cDNA library derived from mRNA in cells of human colorectal and germ cell tumors, and lymphocyte leukemia cells, using the expressed sequence tag (EST) analysis (Adams, M.D., *et al.* Science (1991) 252:1651-1656; Adams, M.D. *et al.*, Nature, (1992) *355*:632-634; Adams, M.D., *et al.*, Nature (1995) 377 Supp:3-174). Polynucleotides of the invention can also be obtained from natural sources such as genomic DNA libraries or can be synthesized using well known and commercially available techniques.

When polynucleotides of the present invention are used for the recombinant production of polypeptides of the present invention, the polynucleotide may include the coding sequence for the mature polypeptide, by itself; or the coding sequence for the mature polypeptide in reading frame with other coding sequences, such as those encoding a leader or secretory sequence, a pre-, or pro- or prepro- protein sequence, or other fusion peptide portions. For example, a marker sequence which facilitates purification of the fused polypeptide can be encoded. In certain preferred embodiments of this aspect of the invention, the marker sequence is a hexa-histidine peptide, as provided in the pQE vector (Qiagen, Inc.) and described in Gentz *et al.*, Proc Natl Acad Sci USA (1989) 86:821-824, or is an HA tag. The polynucleotide may also contain non-coding 5' and 3' sequences, such as transcribed, non-translated sequences, splicing and polyadenylation signals, ribosome binding sites and sequences that stabilize mRNA.

Further embodiments of the present invention include polynucleotides encoding polypeptide variants which comprise the amino acid sequence of SEQ ID NO:2 and SEQ ID NO:3 and in which several, for instance from 5 to 10, 1 to 5, 1 to 3, 1 to 2 or 1, amino acid residues are substituted, deleted or added, in any combination.

Polynucleotides which are identical or sufficiently identical to a nucleotide sequence contained in SEQ ID NO:1, may be used as hybridization probes for cDNA and genomic DNA or as primers for a nucleic acid amplification (PCR) reaction, to isolate full-length cDNAs and genomic clones encoding polypeptides of the present invention and to isolate cDNA and genomic clones of other genes (including genes encoding homologs and orthologs from species other than human) that have a high sequence similarity to SEQ ID NO:1. Typically these nucleotide sequences are 70% identical, preferably 80% identical, more preferably 90% identical, most preferably 95% identical to that of the referent. The probes or primers will generally comprise at least 15 nucleotides, preferably, at least 30 nucleotides and may have at least 50 nucleotides. Particularly preferred probes will have between 30 and 50 nucleotides. In particular, polypeptides or polynucleotides derived from sequences from homologous animal origin could be used as immunogens to obtain a cross-reactive immune response to the human gene.

A polynucleotide encoding a polypeptide of the present invention, including homologs and orthologs from species other than human, may be obtained by a process which comprises the steps of screening an appropriate library under stringent hybridization conditions with a labeled probe having the sequence of SEQ ID NO: 1 or a fragment thereof; and isolating full-length cDNA and genomic clones containing said polynucleotide sequence. Such hybridization techniques are well known to the skilled artisan. Preferred stringent hybridization conditions include overnight incubation at 42°C in a solution comprising: 50% formamide, 5xSSC (150mM NaCl, 15mM trisodium citrate), 50 mM sodium phosphate (pH7.6), 5x Denhardt's solution, 10 % dextran sulfate, and 20 microgram/ml denatured, sheared salmon sperm DNA; followed by washing the filters in 0.1x SSC at about 65°C. Thus the present invention also includes polynucleotides obtainable by screening an appropriate library under stingent hybridization conditions with a labeled probe having the sequence of SEQ ID NO:1 or a fragment thereof.

The skilled artisan will appreciate that, in many cases, an isolated cDNA sequence will be incomplete, in that the region coding for the polypeptide is cut short at the 5' end of the cDNA. This is a consequence of reverse transcriptase, an enzyme with inherently low 'processivity' (a measure of the ability of the enzyme to remain attached to the template during the polymerisation reaction), failing to complete a DNA copy of the mRNA template during 1st strand cDNA synthesis.

There are several methods available and well known to those skilled in the art to obtain full-length cDNAs, or extend short cDNAs, for example those based on the method of Rapid Amplification of cDNA ends (RACE) (see, for example, Frohman et al., PNAS USA 85, 8998-9002, 1988). Recent modifications of the technique, exemplified by the Marathon™ technology (Clontech Laboratories Inc.) for example, have significantly simplified the search for longer cDNAs. In the Marathon™ technology, cDNAs have been prepared from mRNA extracted from a chosen tissue and an 'adaptor' sequence ligated onto each end. Nucleic acid amplification (PCR) is then carried out to amplify the 'missing' 5' end of the cDNA using a combination of gene specific and adaptor specific oligonucleotide primers. The PCR reaction is then repeated using 'nested' primers, that is, primers designed to anneal within the amplified product (typically an adaptor specific primer that anneals further 3' in the adaptor sequence and a gene specific primer that anneals further 5' in the known gene sequence). The products of this reaction can then be analysed by DNA sequencing and a full-length cDNA constructed either by joining the product directly to the existing cDNA to give a complete sequence, or carrying out a separate full-length PCR using the new sequence information for the design of the 5' primer.

Recombinant polypeptides of the present invention may be prepared by processes well known in the art from genetically engineered host cells comprising expression systems. Accordingly, in a further aspect, the present invention relates to expression systems which comprise a polynucleotide or polynucleotides of the present invention, to host cells which are genetically engineered with such expression sytems and to the production of polypeptides of the invention by recombinant techniques. Cell-free translation systems can also be employed to produce such proteins using RNAs derived from the DNA constructs of the present invention.

For recombinant production, host cells can be genetically engineered to incorporate expression systems or portions thereof for polynucleotides of the present invention. Introduction of polynucleotides into host cells can be effected by methods described in many standard laboratory manuals, such as Davis et al., Basic Methods in Molecular Biology (1986) and Sambrook *et al.*, Molecular Cloning: A Laboratory Manual, 2nd Ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. (1989). Preferred such methods include, for instance, calcium phosphate transfection, DEAE-dextran mediated transfection, transvection, microinjection, cationic lipid-mediated transfection, electroporation, transduction, scrape loading, ballistic introduction or infection.

Representative examples of appropriate hosts include bacterial cells, such as *streptococci, staphylococci, E. coli, Streptomyces* and *Bacillus subtilis* cells; fungal cells, such as yeast cells and

Aspergillus cells; insect cells such as *Drosophila* S2 and *Spodoptera* Sf9 cells; animal cells such as CHO, COS, HeLa, C127, 3T3, BHK, HEK 293 and Bowes melanoma cells; and plant cells.

A great variety of expression systems can be used, for instance, chromosomal, episomal and virus-derived systems, e.g., vectors derived from bacterial plasmids, from bacteriophage, from transposons, from yeast episomes, from insertion elements, from yeast chromosomal elements, from viruses such as baculoviruses, papova viruses, such as SV40, vaccinia viruses, adenoviruses, fowl pox viruses, pseudorabies viruses and retroviruses, and vectors derived from combinations thereof, such as those derived from plasmid and bacteriophage genetic elements, such as cosmids and phagemids. The expression systems may contain control regions that regulate as well as engender expression. Generally, any system or vector which is able to maintain, propagate or express a polynucleotide to produce a polypeptide in a host may be used. The appropriate nucleotide sequence may be inserted into an expression system by any of a variety of well-known and routine techniques, such as, for example, those set forth in Sambrook *et al.*, Molecular Cloning, A Laboratory Manual (supra). Appropriate secretion signals may be incorporated into the desired polypeptide to allow secretion of the translated protein into the lumen of the endoplasmic reticulum, the periplasmic space or the extracellular environment. These signals may be endogenous to the polypeptide or they may be heterologous signals.

If a polypeptide of the present invention is to be expressed for use in screening assays, it is generally preferred that the polypeptide be produced at the surface of the cell. In this event, the cells may be harvested prior to use in the screening assay. If the polypeptide is secreted into the medium, the medium can be recovered in order to recover and purify the polypeptide. If produced intracellularly, the cells must first be lysed before the polypeptide is recovered.

Polypeptides of the present invention can be recovered and purified from recombinant cell cultures by well-known methods including ammonium sulfate or ethanol precipitation, acid extraction, anion or cation exchange chromatography, phosphocellulose chromatography, hydrophobic interaction chromatography, affinity chromatography, hydroxylapatite chromatography and lectin chromatography. Most preferably, high performance liquid chromatography is employed for purification. Well known techniques for refolding proteins may be employed to regenerate active conformation when the polypeptide is denatured during isolation and or purification.

Another important aspect of the invention relates to a method for inducing , re-inforcing or modulating an immunological response in a mammal which comprises inoculating the mammal with a fragment or the entire polypeptide or polynucleotide of the present invention, adequate to produce antibody and/or T cell immune response for prophylaxis or for therapeutic treatment of the diseases hereinbefore mentioned, amongst others. Yet another aspect of the invention relates

to a method of inducing, re-inforcing or modulating immunological response in a mammal which comprises, delivering a polypeptide of the present invention *via* a vector or cell directing expression of the polynucleotide and coding for the polypeptide *in vivo* in order to induce such an immunological response to produce immune responses for prophylaxis or treatment of said mammal from diseases.

A further aspect of the invention relates to an immunological/vaccine formulation (composition) which, when introduced into a mammalian host, induces, re-inforces or modulates an immunological response in that mammal to a polypeptide of the present invention wherein the composition comprises a fragment or the entire polypeptide or polynucleotide of the present invention. The vaccine formulation may further comprise a suitable carrier. Since a polypeptide may be broken down in the stomach, it is preferably administered parenterally (for instance, subcutaneous, intramuscular, intravenous, or intradermal injection). Formulations suitable for parenteral administration include aqueous and non-aqueous sterile injection solutions which may contain anti-oxidants, buffers, bacteriostats and solutes which render the formulation isotonic with the blood of the recipient; and aqueous and non-aqueous sterile suspensions which may include suspending agents or thickening agents. The formulations may be presented in unit-dose or multi-dose containers, for example, sealed ampoules and vials and may be stored in a freeze-dried condition requiring only the addition of the sterile liquid carrier immediately prior to use. The vaccine formulation may also include adjuvant systems for enhancing the immunogenicity of the formulation, such as oil-in water systems, immunomodulators and other systems known in the art. The dosage will depend on the specific activity of the vaccine and can be readily determined by routine experimentation.

A further aspect of the invention relates to the *in vitro* induction of immune responses to a fragment or the entire polypeptide or polynucleotide of the present invention or a molecule comprising the polypeptide or polynucleotide of the present invention, using cells from the immune system of a mammal, and reinfusing these activated immune cells of the mammal for the treatment of disease. Activation of the cells from the immune system is achieved by in vitro incubation with the entire polypeptide or polynucleotide of the present invention or a molecule comprising the polypeptide or polynucleotide of the present invention in the presence or absence of various immunomodulator molecules.

A further aspect of the invention relates to the immunization of a mammal by administration of antigen presenting cells modified by *in vitro* loading with part or the entire polypeptide of the present invention or a molecule comprising the polypeptide of the present invention and administered *in vivo* in an immunogenic way. Alternatively, antigen presenting cells can be transfected *in vitro* with a vector containing a fragment or the entire polynucleotide of

the present invention or a molecule comprising the polynucleotide of the present invention, such as to express the corresponding polypeptide, and administered *in vivo* in an immunogenic way.

This invention also relates to the use of polynucleotides of the present invention as diagnostic reagents. Detection of a mutated form of the gene characterised by the polynucleotide of SEQ ID NO:1 which is associated with a dysfunction will provide a diagnostic tool that can add to, or define, a diagnosis of a disease, or susceptibility to a disease, which results from under-expression, over-expression or altered expression of the gene. Individuals carrying mutations in the gene may be detected at the DNA level by a variety of techniques.

Nucleic acids for diagnosis may be obtained from a subject's cells, such as from blood, urine, saliva, tissue biopsy or autopsy material. The genomic DNA may be used directly for detection or may be amplified enzymatically by using PCR or other amplification techniques prior to analysis. RNA or cDNA may also be used in similar fashion. Deletions and insertions can be detected by a change in size of the amplified product in comparison to the normal genotype. Point mutations can be identified by hybridizing amplified DNA to labeled CASB7439 nucleotide sequences. Perfectly matched sequences can be distinguished from mismatched duplexes by RNase digestion or by differences in melting temperatures. DNA sequence differences may also be detected by alterations in electrophoretic mobility of DNA fragments in gels, with or without denaturing agents, or by direct DNA sequencing (ee, e.g., Myers *et al.*, Science (1985) 230:1242). Sequence changes at specific locations may also be revealed by nuclease protection assays, such as RNase and S1 protection or the chemical cleavage method (see Cotton *et al.*, Proc Natl Acad Sci USA (1985) 85: 4397-4401). In another embodiment, an array of oligonucleotides probes comprising CASB7439 nucleotide sequence or fragments thereof can be constructed to conduct efficient screening of e.g., genetic mutations. Array technology methods are well known and have general applicability and can be used to address a variety of questions in molecular genetics including gene expression, genetic linkage, and genetic variability (see for example: M.Chee et al., Science, Vol 274, pp 610-613 (1996)).

The diagnostic assays offer a process for diagnosing or determining a susceptibility to the Diseases through detection of mutation in the CASB7439 gene by the methods described. In addition, such diseases may be diagnosed by methods comprising determining from a sample derived from a subject an abnormally decreased or increased level of polypeptide or mRNA. Decreased or increased expression can be measured at the RNA level using any of the methods well known in the art for the quantitation of polynucleotides, such as, for example, nucleic acid amplification, for instance PCR, RT-PCR, RNase protection, Northern blotting and other hybridization methods. Assay techniques that can be used to determine levels of a protein, such as a polypeptide of the present invention, in a sample derived from a host are well-known to those of skill

in the art. Such assay methods include radioimmunoassays, competitive-binding assays, Western Blot analysis and ELISA assays.

Thus in another aspect, the present invention relates to a diagnostic kit which comprises:
(a) a polynucleotide of the present invention, preferably the nucleotide sequence of SEQ ID NO: 1, or a fragment thereof;
(b) a nucleotide sequence complementary to that of (a);
(c) a polypeptide of the present invention, preferably the polypeptide of SEQ ID NO:2 and SEQ ID NO:3 or a fragment thereof; or
(d) an antibody to a polypeptide of the present invention, preferably to the polypeptide of SEQ ID NO:2 and SEQ ID NO:3.

It will be appreciated that in any such kit, (a), (b), (c) or (d) may comprise a substantial component. Such a kit will be of use in diagnosing a disease or suspectability to a disease, particularly, amongst others.

The nucleotide sequences of the present invention are also valuable for chromosome identification. The sequence is specifically targeted to, and can hybridize with, a particular location on an individual human chromosome. The mapping of relevant sequences to chromosomes according to the present invention is an important first step in correlating those sequences with gene associated disease. Once a sequence has been mapped to a precise chromosomal location, the physical position of the sequence on the chromosome can be correlated with genetic map data. Such data are found in, for example, V. McKusick, Mendelian Inheritance in Man (available on-line through Johns Hopkins University Welch Medical Library). The relationship between genes and diseases that have been mapped to the same chromosomal region are then identified through linkage analysis (coinheritance of physically adjacent genes).

The differences in the cDNA or genomic sequence between affected and unaffected individuals can also be determined. If a mutation is observed in some or all of the affected individuals but not in any normal individuals, then the mutation is likely to be the causative agent of the disease.

The polypeptides of the invention or their fragments or analogs thereof, or cells expressing them, can also be used as immunogens to produce antibodies immunospecific for polypeptides of the present invention. The term "immunospecific" means that the antibodies have substantially greater affinity for the polypeptides of the invention than their affinity for other related polypeptides in the prior art.

Antibodies generated against polypeptides of the present invention may be obtained by administering the polypeptides or epitope-bearing fragments, analogs or cells to an animal, preferably a non-human animal, using routine protocols. For preparation of monoclonal antibodies, any technique which provides antibodies produced by continuous cell line cultures can be used. Examples include the hybridoma technique (Kohler, G. and Milstein, C., Nature (1975) 256:495-497), the trioma technique, the human B-cell hybridoma technique (Kozbor *et al.*, Immunology Today (1983) 4:72) and the EBV-hybridoma technique (Cole *et al.*, Monoclonal Antibodies and Cancer Therapy, 77-96, Alan R. Liss, Inc., 1985).

Techniques for the production of single chain antibodies, such as those described in U.S. Patent No. 4,946,778, can also be adapted to produce single chain antibodies to polypeptides of this invention. Also, transgenic mice, or other organisms, including other mammals, may be used to express humanized antibodies.

The above-described antibodies may be employed to isolate or to identify clones expressing the polypeptide or to purify the polypeptides by affinity chromatography.

Antibodies against polypeptides of the present invention may also be employed to treat the Diseases, amongst others.

In a further aspect, the present invention relates to genetically engineered soluble fusion proteins comprising a polypeptide of the present invention, or a fragment thereof, and various portions of the constant regions of heavy or light chains of immunoglobulins of various subclasses (IgG, IgM, IgA, IgE). Preferred as an immunoglobulin is the constant part of the heavy chain of human IgG, particularly IgG1, where fusion takes place at the hinge region. In a particular embodiment, the Fc part can be removed simply by incorporation of a cleavage sequence which can be cleaved with blood clotting factor Xa. Furthermore, this invention relates to processes for the preparation of these fusion proteins by genetic engineering, and to the use thereof for drug screening, diagnosis and therapy. A further aspect of the invention also relates to polynucleotides encoding such fusion proteins. Examples of fusion protein technology can be found in International Patent Application Nos. WO94/29458 and WO94/22914.

Another aspect of the invention relates to a method for inducing an immunological response in a mammal which comprises inoculating the mammal with a polypeptide of the present invention, adequate to produce antibody and/or T cell immune response to protect said animal from the Diseases hereinbefore mentioned, amongst others. Yet another aspect of the invention relates to a method of inducing immunological response in a mammal which comprises, delivering a polypeptide of the present invention *via* a vector directing expression of the

polynucleotide and coding for the polypeptide *in vivo* in order to induce such an immunological response to produce antibody to protect said animal from diseases.

Polypeptides of the present invention are responsible for many biological functions, including many disease states, in particular the Diseases hereinbefore mentioned. It is therefore desirous to devise screening methods to identify compounds which stimulate or which inhibit the function of the polypeptide. Accordingly, in a further aspect, the present invention provides for a method of screening compounds to identify those which stimulate or which inhibit the function of the polypeptide. In general, agonists or antagonists may be employed for therapeutic and prophylactic purposes for such Diseases as hereinbefore mentioned. Compounds may be identified from a variety of sources, for example, cells, cell-free preparations, chemical libraries, and natural product mixtures. Such agonists, antagonists or inhibitors so-identified may be natural or modified substrates, ligands, receptors, enzymes, etc., as the case may be, of the polypeptide; or may be structural or functional mimetics thereof (see Coligan *et al.*, Current Protocols in Immunology 1(2):Chapter 5 (1991)).

The screening method may simply measure the binding of a candidate compound to the polypeptide, or to cells or membranes bearing the polypeptide, or a fusion protein thereof by means of a label directly or indirectly associated with the candidate compound. Alternatively, the screening method may involve competition with a labeled competitor. Further, these screening methods may test whether the candidate compound results in a signal generated by activation or inhibition of the polypeptide, using detection systems appropriate to the cells bearing the polypeptide. Inhibitors of activation are generally assayed in the presence of a known agonist and the effect on activation by the agonist by the presence of the candidate compound is observed. Constitutively active polpypeptides may be employed in screening methods for inverse agonists or inhibitors, in the absence of an agonist or inhibitor, by testing whether the candidate compound results in inhibition of activation of the polypeptide. Further, the screening methods may simply comprise the steps of mixing a candidate compound with a solution containing a polypeptide of the present invention, to form a mixture, measuring CASB7439 activity in the mixture, and comparing the CASB7439 activity of the mixture to a standard. Fusion proteins, such as those made from Fc portion and CASB7439 polypeptide, as hereinbefore described, can also be used for high-throughput screening assays to identify antagonists for the polypeptide of the present invention (see D. Bennett *et al.*, J Mol Recognition, 8:52-58 (1995); and K. Johanson *et al.*, J Biol Chem, 270(16):9459-9471 (1995)).

The polynucleotides, polypeptides and antibodies to the polypeptide of the present invention may also be used to configure screening methods for detecting the effect of added

compounds on the production of mRNA and polypeptide in cells. For example, an ELISA assay may be constructed for measuring secreted or cell associated levels of polypeptide using monoclonal and polyclonal antibodies by standard methods known in the art. This can be used to discover agents which may inhibit or enhance the production of polypeptide (also called antagonist or agonist, respectively) from suitably manipulated cells or tissues.

The polypeptide may be used to identify membrane bound or soluble receptors, if any, through standard receptor binding techniques known in the art. These include, but are not limited to, ligand binding and crosslinking assays in which the polypeptide is labeled with a radioactive isotope (for instance, ^{125}I), chemically modified (for instance, biotinylated), or fused to a peptide sequence suitable for detection or purification, and incubated with a source of the putative receptor (cells, cell membranes, cell supernatants, tissue extracts, bodily fluids). Other methods include biophysical techniques such as surface plasmon resonance and spectroscopy. These screening methods may also be used to identify agonists and antagonists of the polypeptide which compete with the binding of the polypeptide to its receptors, if any. Standard methods for conducting such assays are well understood in the art.

Examples of potential polypeptide antagonists include antibodies or, in some cases, oligonucleotides or proteins which are closely related to the ligands, substrates, receptors, enzymes, etc., as the case may be, of the polypeptide, e.g., a fragment of the ligands, substrates, receptors, enzymes, etc.; or small molecules which bind to the polypetide of the present invention but do not elicit a response, so that the activity of the polypeptide is prevented.

Thus, in another aspect, the present invention relates to a screening kit for identifying agonists, antagonists, ligands, receptors, substrates, enzymes, etc. for polypeptides of the present invention; or compounds which decrease or enhance the production of such polypeptides, which comprises:

(a) a polypeptide of the present invention;

(b) a recombinant cell expressing a polypeptide of the present invention;

(c) a cell membrane expressing a polypeptide of the present invention; or

(d) antibody to a polypeptide of the present invention;

which polypeptide is preferably that of SEQ ID NO:2 and SEQ ID NO:3.

It will be appreciated that in any such kit, (a), (b), (c) or (d) may comprise a substantial component.

It will be readily appreciated by the skilled artisan that a polypeptide of the present invention may also be used in a method for the structure-based design of an agonist, antagonist or inhibitor of the polypeptide, by:

(a) determining in the first instance the three-dimensional structure of the polypeptide;

(b) deducing the three-dimensional structure for the likely reactive or binding site(s) of an agonist, antagonist or inhibitor;

(c) synthesing candidate compounds that are predicted to bind to or react with the deduced binding or reactive site; and

(d) testing whether the candidate compounds are indeed agonists, antagonists or inhibitors.

It will be further appreciated that this will normally be an interative process.

In a further aspect, the present invention provides methods of treating abnormal conditions such as, for instance, cancer and autoimmune diseases, related to either an excess of, or an under-expression of, CASB7439 polypeptide activity.

If the activity of the polypeptide is in excess, several approaches are available. One approach comprises administering to a subject in need thereof an inhibitor compound (antagonist) as hereinabove described, optionally in combination with a pharmaceutically acceptable carrier, in an amount effective to inhibit the function of the polypeptide, such as, for example, by blocking the binding of ligands, substrates, receptors, enzymes, etc., or by inhibiting a second signal, and thereby alleviating the abnormal condition. In another approach, soluble forms of the polypeptides still capable of binding the ligand, substrate, enzymes, receptors, etc. in competition with endogenous polypeptide may be administered. Typical examples of such competitors include fragments of the CASB7439 polypeptide.

In still another approach, expression of the gene encoding endogenous CASB7439 polypeptide can be inhibited using expression blocking techniques. Known such techniques involve the use of antisense sequences, either internally generated or separately administered (see, for example, O'Connor, J Neurochem (1991) 56:560 in Oligodeoxynucleotides as Antisense Inhibitors of Gene Expression, CRC Press, Boca Raton, FL (1988)). Alternatively, oligonucleotides which form triple helices with the gene can be supplied (see, for example, Lee *et al.*, Nucleic Acids Res (1979) 6:3073; Cooney *et al.*, Science (1988) 241:456; Dervan *et al.*, Science (1991) 251:1360). These oligomers can be administered *per se* or the relevant oligomers can be expressed *in vivo*.

For treating abnormal conditions related to an under-expression of CASB7439 and its activity, several approaches are also available. One approach comprises administering to a subject a therapeutically effective amount of a compound which activates a polypeptide of the present invention, i.e., an agonist as described above, in combination with a pharmaceutically acceptable carrier, to thereby alleviate the abnormal condition. Alternatively, gene therapy may be employed to effect the endogenous production of CASB7439 by the relevant cells in the subject. For example, a

polynucleotide of the invention may be engineered for expression in a replication defective retroviral vector, as discussed above. The retroviral expression construct may then be isolated and introduced into a packaging cell transduced with a retroviral plasmid vector containing RNA encoding a polypeptide of the present invention such that the packaging cell now produces infectious viral particles containing the gene of interest. These producer cells may be administered to a subject for engineering cells *in vivo* and expression of the polypeptide *in vivo*. For an overview of gene therapy, see Chapter 20, Gene Therapy and other Molecular Genetic-based Therapeutic Approaches, (and references cited therein) in Human Molecular Genetics, T Strachan and A P Read, BIOS Scientific Publishers Ltd (1996). Another approach is to administer a therapeutic amount of a polypeptide of the present invention in combination with a suitable pharmaceutical carrier.

In a further aspect, the present invention provides for pharmaceutical compositions comprising a therapeutically effective amount of a polypeptide, such as the soluble form of a polypeptide of the present invention, agonist/antagonist peptide or small molecule compound, in combination with a pharmaceutically acceptable carrier or excipient. Such carriers include, but are not limited to, saline, buffered saline, dextrose, water, glycerol, ethanol, and combinations thereof. The invention further relates to pharmaceutical packs and kits comprising one or more containers filled with one or more of the ingredients of the aforementioned compositions of the invention. Polypeptides and other compounds of the present invention may be employed alone or in conjunction with other compounds, such as therapeutic compounds.

The composition will be adapted to the route of administration, for instance by a systemic or an oral route. Preferred forms of systemic administration include injection, typically by intravenous injection. Other injection routes, such as subcutaneous, intramuscular, or intraperitoneal, can be used. Alternative means for systemic administration include transmucosal and transdermal administration using penetrants such as bile salts or fusidic acids or other detergents. In addition, if a polypeptide or other compounds of the present invention can be formulated in an enteric or an encapsulated formulation, oral administration may also be possible. Administration of these compounds may also be topical and/or localized, in the form of salves, pastes, gels, and the like.

The dosage range required depends on the choice of peptide or other compounds of the present invention, the route of administration, the nature of the formulation, the nature of the subject's condition, and the judgment of the attending practitioner. Suitable dosages, however, are in the range of 0.1-100 µg/kg of subject. Wide variations in the needed dosage, however, are to be expected in view of the variety of compounds available and the differing efficiencies of various routes of administration. For example, oral administration would be expected to require higher

dosages than administration by intravenous injection. Variations in these dosage levels can be adjusted using standard empirical routines for optimization, as is well understood in the art.

Polypeptides used in treatment can also be generated endogenously in the subject, in treatment modalities often referred to as "gene therapy" as described above. Thus, for example, cells from a subject may be engineered with a polynucleotide, such as a DNA or RNA, to encode a polypeptide *ex vivo*, and for example, by the use of a retroviral plasmid vector. The cells are then introduced into the subject.

Polynucleotide and polypeptide sequences form a valuable information resource with which to identify further sequences of similar homology. This is most easily facilitated by storing the sequence in a computer readable medium and then using the stored data to search a sequence database using well known searching tools, such as GCC. Accordingly, in a further aspect, the present invention provides for a computer readable medium having stored thereon a polynucleotide comprising the sequence of SEQ ID NO:1 and/or a polypeptide sequence encoded thereby.

The following definitions are provided to facilitate understanding of certain terms used frequently hereinbefore.

"Antibodies" as used herein includes polyclonal and monoclonal antibodies, chimeric, single chain, and humanized antibodies, as well as Fab fragments, including the products of an Fab or other immunoglobulin expression library.

"Isolated" means altered "by the hand of man" from the natural state. If an "isolated" composition or substance occurs in nature, it has been changed or removed from its original environment, or both. For example, a polynucleotide or a polypeptide naturally present in a living animal is not "isolated," but the same polynucleotide or polypeptide separated from the coexisting materials of its natural state is "isolated", as the term is employed herein.

"Polynucleotide" generally refers to any polyribonucleotide or polydeoxribonucleotide, which may be unmodified RNA or DNA or modified RNA or DNA. "Polynucleotides" include, without limitation, single- and double-stranded DNA, DNA that is a mixture of single- and double-stranded regions, single- and double-stranded RNA, and RNA that is mixture of single- and double-stranded regions, hybrid molecules comprising DNA and RNA that may be single-stranded or, more typically, double-stranded or a mixture of single- and double-stranded regions. In addition, "polynucleotide" refers to triple-stranded regions comprising RNA or DNA or both RNA and DNA. The term "polynucleotide" also includes DNAs or RNAs containing one or more modified bases and DNAs or RNAs with backbones modified for stability or for other reasons. "Modified" bases include, for example, tritylated bases and unusual bases such as inosine. A

variety of modifications may be made to DNA and RNA; thus, "polynucleotide" embraces chemically, enzymatically or metabolically modified forms of polynucleotides as typically found in nature, as well as the chemical forms of DNA and RNA characteristic of viruses and cells. "Polynucleotide" also embraces relatively short polynucleotides, often referred to as oligonucleotides.

"Polypeptide" refers to any peptide or protein comprising two or more amino acids joined to each other by peptide bonds or modified peptide bonds, i.e., peptide isosteres. "Polypeptide" refers to both short chains, commonly referred to as peptides, oligopeptides or oligomers, and to longer chains, generally referred to as proteins. Polypeptides may contain amino acids other than the 20 gene-encoded amino acids. "Polypeptides" include amino acid sequences modified either by natural processes, such as post-translational processing, or by chemical modification techniques which are well known in the art. Such modifications are well described in basic texts and in more detailed monographs, as well as in a voluminous research literature. Modifications may occur anywhere in a polypeptide, including the peptide backbone, the amino acid side-chains and the amino or carboxyl termini. It will be appreciated that the same type of modification may be present to the same or varying degrees at several sites in a given polypeptide. Also, a given polypeptide may contain many types of modifications. Polypeptides may be branched as a result of ubiquitination, and they may be cyclic, with or without branching. Cyclic, branched and branched cyclic polypeptides may result from post-translation natural processes or may be made by synthetic methods. Modifications include acetylation, acylation, ADP-ribosylation, amidation, covalent attachment of flavin, covalent attachment of a heme moiety, covalent attachment of a nucleotide or nucleotide derivative, covalent attachment of a lipid or lipid derivative, covalent attachment of phosphotidylinositol, cross-linking, cyclization, disulfide bond formation, demethylation, formation of covalent cross-links, formation of cystine, formation of pyroglutamate, formylation, gamma-carboxylation, glycosylation, GPI anchor formation, hydroxylation, iodination, methylation, myristoylation, oxidation, proteolytic processing, phosphorylation, prenylation, racemization, selenoylation, sulfation, transfer-RNA mediated addition of amino acids to proteins such as arginylation, and ubiquitination (see, for instance, Proteins - Structure and Molecular Properties, 2nd Ed., T. E. Creighton, W. H. Freeman and Company, New York, 1993; Wold, F., Post-translational Protein Modifications: Perspectives and Prospects, pgs. 1-12 in Post-translational Covalent Modification of Proteins, B. C. Johnson, Ed., Academic Press, New York, 1983; Seifter *et al.*, "Analysis for protein modifications and nonprotein cofactors", Meth Enzymol (1990) 182:626-646 and Rattan *et al.*, "Protein Synthesis: Post-translational Modifications and Aging", Ann NY Acad Sci (1992) 663:48-62).

"Variant" refers to a polynucleotide or polypeptide that differs from a reference polynucleotide or polypeptide, but retains essential properties. A typical variant of a polynucleotide differs in nucleotide sequence from another, reference polynucleotide. Changes in the nucleotide sequence of the variant may or may not alter the amino acid sequence of a polypeptide encoded by the reference polynucleotide. Nucleotide changes may result in amino acid substitutions, additions, deletions, fusions and truncations in the polypeptide encoded by the reference sequence, as discussed below. A typical variant of a polypeptide differs in amino acid sequence from another, reference polypeptide. Generally, differences are limited so that the sequences of the reference polypeptide and the variant are closely similar overall and, in many regions, identical. A variant and reference polypeptide may differ in amino acid sequence by one or more substitutions, additions, deletions in any combination. A substituted or inserted amino acid residue may or may not be one encoded by the genetic code. A variant of a polynucleotide or polypeptide may be a naturally occurring such as an allelic variant, or it may be a variant that is not known to occur naturally. Non-naturally occurring variants of polynucleotides and polypeptides may be made by mutagenesis techniques or by direct synthesis.

"Identity," as known in the art, is a relationship between two or more polypeptide sequences or two or more polynucleotide sequences, as determined by comparing the sequences. In the art, "identity" also means the degree of sequence relatedness between polypeptide or polynucleotide sequences, as the case may be, as determined by the match between strings of such sequences. "Identity" and "similarity" can be readily calculated by known methods, including but not limited to those described in (Computational Molecular Biology, Lesk, A.M., ed., Oxford University Press, New York, 1988; Biocomputing: Informatics and Genome Projects, Smith, D.W., ed., Academic Press, New York, 1993; Computer Analysis of Sequence Data, Part I, Griffin, A.M., and Griffin, H.G., eds., Humana Press, New Jersey, 1994; Sequence Analysis in Molecular Biology, von Heinje, G., Academic Press, 1987; and Sequence Analysis Primer, Gribskov, M. and Devereux, J., eds., M Stockton Press, New York, 1991; and Carillo, H., and Lipman, D., SIAM *J. Applied Math., 48:* 1073 (1988). Preferred methods to determine identity are designed to give the largest match between the sequences tested. Methods to determine identity and similarity are codified in publicly available computer programs. Preferred computer program methods to determine identity and similarity between two sequences include, but are not limited to, the GCG program package (Devereux, J., et al., Nucleic Acids Research *12(1):* 387 (1984)), BLASTP, BLASTN, and FASTA (Atschul, S.F. et al., J. Molec. Biol. *215:* 403-410 (1990). The BLAST X program is publicly available from NCBI and other sources (BLAST Manual, Altschul, S., *et al.*, NCBI

NLM NIH Bethesda, MD 20894; Altschul, S., *et al.*, J. Mol. Biol. *215*: 403-410 (1990). The well known Smith Waterman algorithm may also be used to determine identity.

Preferred parameters for polypeptide sequence comparison include the following:

1) Algorithm: Needleman and Wunsch, J. Mol Biol. 48: 443-453 (1970)

Comparison matrix: BLOSSUM62 from Hentikoff and Hentikoff, Proc. Natl. Acad. Sci. USA. 89:10915-10919 (1992)

Gap Penalty: 12

Gap Length Penalty: 4

A program useful with these parameters is publicly available as the "gap" program from Genetics Computer Group, Madison WI. The aforementioned parameters are the default parameters for polypeptide comparisons (along with no penalty for end gaps).

Preferred parameters for polynucleotide comparison include the following:

1) Algorithm: Needleman and Wunsch, J. Mol Biol. 48: 443-453 (1970)

Comparison matrix: matches = +10, mismatch = 0

Gap Penalty: 50

Gap Length Penalty: 3

A program useful with these parameters is publicly available as the "gap" program from Genetics Computer Group, Madison WI. The aforementioned parameters are the default parameters for polynucleotide comparisons.

By way of example, a polynucleotide sequence of the present invention may be identical to the reference sequence of SEQ ID NO:1, that is be 100% identical, or it may include up to a certain integer number of nucleotide alterations as compared to the reference sequence. Such alterations are selected from the group consisting of at least one nucleotide deletion, substitution, including transition and transversion, or insertion, and wherein said alterations may occur at the 5' or 3' terminal positions of the reference nucleotide sequence or anywhere between those terminal positions, interspersed either individually among the nucleotides in the reference sequence or in one or more contiguous groups within the reference sequence. The number of nucleotide alterations is determined by multiplying the total number of nucleotides in SEQ ID NO:1 by the numerical percent of the respective percent identity(divided by 100) and subtracting that product from said total number of nucleotides in SEQ ID NO:1, or:

$$n_n \leq x_n - (x_n \bullet y),$$

wherein n_n is the number of nucleotide alterations, x_n is the total number of nucleotides in SEQ ID NO:1, and y is, for instance, 0.70 for 70%, 0.80 for 80%, 0.85 for 85%, 0.90 for 90%, 0.95 for 95%,etc., and wherein any non-integer product of x_n and y is rounded down to the nearest integer

prior to subtracting it from x_n. Alterations of a polynucleotide sequence encoding the polypeptide of SEQ ID NO:2 and SEQ ID NO:3 may create nonsense, missense or frameshift mutations in this coding sequence and thereby alter the polypeptide encoded by the polynucleotide following such alterations.

Similarly, a polypeptide sequence of the present invention may be identical to the reference sequence of SEQ ID NO:2 and SEQ ID NO:3, that is be 100% identical, or it may include up to a certain integer number of amino acid alterations as compared to the reference sequence such that the % identity is less than 100%. Such alterations are selected from the group consisting of at least one amino acid deletion, substitution, including conservative and non-conservative substitution, or insertion, and wherein said alterations may occur at the amino- or carboxy-terminal positions of the reference polypeptide sequence or anywhere between those terminal positions, interspersed either individually among the amino acids in the reference sequence or in one or more contiguous groups within the reference sequence. The number of amino acid alterations for a given % identity is determined by multiplying the total number of amino acids in SEQ ID NO:2 and SEQ ID NO:3 by the numerical percent of the respective percent identity(divided by 100) and then subtracting that product from said total number of amino acids in SEQ ID NO:2 and SEQ ID NO:3, or:

$$n_a \leq x_a - (x_a \bullet y),$$

wherein n_a is the number of amino acid alterations, x_a is the total number of amino acids in SEQ ID NO:2 and SEQ ID NO:3, and y is, for instance 0.70 for 70%, 0.80 for 80%, 0.85 for 85% etc., and wherein any non-integer product of x_a and y is rounded down to the nearest integer prior to subtracting it from x_a.

"Fusion protein" refers to a protein encoded by two, often unrelated, fused genes or fragments thereof. In one example, EP-A-0 464 discloses fusion proteins comprising various portions of constant region of immunoglobulin molecules together with another human protein or part thereof. In many cases, employing an immunoglobulin Fc region as a part of a fusion protein is advantageous for use in therapy and diagnosis resulting in, for example, improved pharmacokinetic properties [see, e.g., EP-A 0232 262]. On the other hand, for some uses it would be desirable to be able to delete the Fc part after the fusion protein has been expressed, detected and purified.

All publications, including but not limited to patents and patent applications, cited in this specification are herein incorporated by reference as if each individual publication were

specifically and individually indicated to be incorporated by reference herein as though fully set forth.

Examples

Example 1

Real-time RT-PCR analysis

Real-time RT-PCR (U. Gibson. 1996. Genome Research: 6,996) is used to compare mRNA transcript abundance of the candidate antigen in matched tumour and normal colon tissues from multiple patients. In addition, mRNA levels of the candidate gene in a panel of normal tissues are also evaluated by this approach.

Total RNA from normal and tumour colon is extracted from snap frozen biopsies using TriPure reagent (Boehringer). Total RNA from normal tissues is purchased from InVitrogen or is extracted from snap frozen biopsies using TriPure reagent. Poly-A+ mRNA is purified from total RNA after DNAase treatment using oligo-dT magnetic beads (Dynal). Quantification of the mRNA is performed by spectrofluorimetry (VersaFluor, BioRad) using SybrII dye (Molecular Probes). Primers for real-time PCR amplification are designed with the Perkin-Elmer Primer Express software using default options for TaqMan amplification conditions.

Real-time reactions are assembled according to standard PCR protocols using 2 ng of purified mRNA for each reaction. SybrI dye (Molecular Probes) is added at a final dilution of 1/75000 for real-time detection. Amplification (40 cycles) and real-time detection is performed in a Perkin-Elmer Biosystems PE7700 system using conventional instrument settings. Ct values are calculated using the PE7700 Sequence Detector software. Several Ct values are obtained for each samples : for the patient samples, the tumour Ct (CtT) and the matched normal colon Ct (CtN) values on the candidate TAA, and for the panel of normal tissue samples, a CtXY for each normal tissue XY. An

another Ct (CtA) is also calculated on Actin gene, as an internal reference, for all of the samples.

As the efficiency of PCR amplification under the prevailing experimental conditions is close to the theoretical amplification efficiency, $2^{(CtN/T/CY-CtA)}$ value is an estimate of the relative TAA transcript level of the sample, standardised with respect to Actin transcript level. A value of 1 thus suggests the candidate antigen and Actin have the same expression level.

Real-time PCR reactions were first performed on tumour colon and matching normal colon from biopsies of 12 patients. Reactions were then performed on a more complete data set totalling 18 patients (are included in this data set the first 12 patients). Duplicates for 6 out of these 18 patients were made in this data set. Results are shown in figure 1.

A series of 48 normal tissue samples, representing 29 different tissues, were also tested by the same procedure (analysed normal tissues are given in table 3). TAA transcript levels are calculated as described above. The proportion of patients over-expressing the candidate antigen, as well as the average transcript over-expression versus normal tissues is also calculated from this data set. Results are shown in figure 2.

Overall results are shown in table 1 and table 2 :

Table 1 : CASB7439 Real-time PCR expression results : data set of 12 patients.

% of patients with a mRNA level higher in matched tumour colon (positive patients)	**92%**
% of patients with a mRNA level at least 3 fold higher in matched tumour colon	**92%**
% of patients with a mRNA level at least 10 fold higher in matched tumour colon	**92%**
% of patients with a mRNA level at least 3 fold lower in matched tumour colon.	**8%**
Average matched normal colon mRNA level (Actin standardised)	**0.0026**
Average matched tumour colon mRNA level in positive patients (Actin standardised)	**0.265**
Average mRNA over-expression fold	**2028**
Median mRNA over-expression fold	**115**
Average normal tissues mRNA level	**0.0079**
Median normal tissues mRNA level	**0.0016**
Average non-dispensable normal tissues mRNA level	**0.0064**

Median non-dispensable normal tissues mRNA level	0.0017
% of patients with a mRNA level higher than average non-dispensable normal tissues	92%
% of patients with a mRNA level higher than 10 fold average non-dispensable normal tissues	75%
Normal non-dispensable tissues higher than median non-dispensable normal tissue mRNA level	None

Table 2 : CASB7439 Real-time PCR expression results : data set of 18 patients.

% of patients with a mRNA level higher in matched tumour colon (positive patients)	89%
% of patients with a mRNA level at least 3 fold higher in matched tumour colon	89%
% of patients with a mRNA level at least 10 fold higher in matched tumour colon	78%
% of patients with a mRNA level at least 3 fold lower in matched tumour colon.	5%
Average matched normal colon mRNA level (Actin standardised)	0.005
Average matched tumour colon mRNA level in positive patients (Actin standardised)	0.152
Average mRNA over-expression fold	1100
Median mRNA over-expression fold	60
Average normal tissues mRNA level	0.0065
Median normal tissues mRNA level	0.0015
Average non-dispensable normal tissues mRNA level	0.005
Median non-dispensable normal tissues mRNA level	0.0015
% of patients with a mRNA level higher than median non-dispensable normal tissues	94%
% of patients with a mRNA level higher than 10 fold median non-dispensable normal tissues	94%
Normal non-dispensable tissues higher than median non-dispensable normal tissue mRNA level	None

Table 1 and 2 clearly suggest CASB7439 transcript is over-expressed in colorectal tumours compared to adjacent normal colon and to all of the above mentioned normal tissues. Around nine patients out of ten strongly over-express CASB7439 transcript, with an average over-expression fold of at least a thousand.

Table 3 : listing of normal tissues used for CASB7439 transcript expression analysis.

Tissue	Abbreviation	Categorie
adrenal gland	Ad_Gl	non dispensable
Aorta	Ao	non dispensable
Bladder	Bl	non dispensable
bone marrow	Bo_Ma	non dispensable
Brain	Bra	non dispensable
Cervix	Ce	non dispensable
Colon	Co	non dispensable
fallopian tube	Fa_Tu	non dispensable
Heart	He	non dispensable
Ileon	Il	non dispensable
Kidney	Ki	non dispensable
Liver	Li	non dispensable
Lung	Lu	non dispensable
lymph node	Ly_No	non dispensable
Oesophagus	Oe	non dispensable
Parathyroid gland	Pa_Thy	non dispensable
Rectum	Re	non dispensable
Skin	Sk	non dispensable
skeletal muscle	Sk_Mu	non dispensable
small intestine	Sm_In	non dispensable
Spleen	Sp	non dispensable
Stomach	St	non dispensable
thyroid gland	Thy	non dispensable
Trachea	Tra	non dispensable
Ovary	Ov	dispensable
Placenta	Pl	dispensable
Prostate	Pr	dispensable
Testis	Te	dispensable

Example 2.

DNA microarrays

DNA micro-arrays are used to examine mRNA expression profiles of large collections of genes in multiple samples. This information is used to complement the data obtained by real-time PCR and provides an independent measure of gene expression levels in tumors and normal tissues.

Examples of current technologies for production of DNA micro-arrays include 1) The Affymetrix "GeneChip" arrays in which oligonucleotides are synthetized on the surface of the chip by solid phase chemical synthesis using a photolithographic process 2) DNA spotting technology in which small volumes of a DNA solution are robotically deposited and then immobilized onto the surface of a solid phase (e.g. glass). In both instances, the chips are hybridized with cDNA or cRNA which has been extracted from the tissue of interest (e.g. normal tissue, tumour etc...) and labeled with radioactivity or with a fluorescent reporter molecule. The labeled material is hybridized to the chip and the amount of probe bound to each sequence on the chip is determined using a specialized scanner. The experiment can be set-up with a single fluorescent reporter (or radioactivity) or, alternatively, can be performed using two fluorescent reporters. In this latter case, each of the two samples is labeled with one of the reporter molecules. The two labeled samples are then hybridized competitively to the sequences on the DNA chip. The ratio of the two fluorescent signals is determined for each sequence on the chip. This ratio is used to calculate the relative abundance of the transcript in the two samples. Detailed protocols are available from a number of sources including "DNA Microarrays: A practical approach. Schena M. Oxford University Press 1999" and the World Wide Web (http://cmgm.stanford.edu/pbrown/protocols/index.html), http://arrayit.com/DNA-Microarray-Protocols/) and specialized distributors (e.g. Affymetrix).

Example 3

Northern-Southern blot analysis

Limited amounts of mixed tumour and matched normal colon cDNA are amplified by Advantage PCR (see above). Messenger RNA from multiple normal tissues is also amplified using the same procedure. The amplified cDNA (1 µg) is electrophoresed on a

1.2% agarose gel and transferred onto a nylon membrane. The membrane is hybridised (AlkPhos Direct System) with a probe prepared using a fragment of the candidate TAA cDNA. Northern-Southern analysis provides information on transcript size, presence of splice variants and transcript abundance in tumour and normal tissues.

Example 4

Northern Blot Analysis

Northern blots are produced according to standard protocols using 1 µg of poly A+ mRNA. Radioactive probes are prepared using the Ready-to-Go system (Pharmacia).

Example 5

Identification of the full length cDNA sequence

Colon tumour cDNA libraries are constructed using the Lambda Zap II system (Stratagene) from 5 µg of polyA+ mRNA. The supplied protocol is followed except that SuperscriptII (Life Technologies) is used for the reverse transcription step. Oligo dT-primed and random-primed libraries are constructed. About 1.5 x106 independent phages are plated for each screening of the library. Phage plaques are transferred onto nylon filters and hybridised using a cDNA probe labelled with AlkPhos Direct. Positive phages are detected by chemiluminescence. Positive phage are excised from the agar plat, eluted in 500µl SM buffer and confirmed by gene-specific PCR. Eluted phages are converted to single strand M13 bacteriophage by in vivo excision. The bacteriophage is then converted to double strand plasmid DNA by infection of E. coli. Infected bacteria are plated and submitted to a second round of screening with the cDNA probe. Plasmid DNA is purified from positive bacterial clones and sequenced on both strands.

When the full length gene cannot be obtained directly from the cDNA library, missing sequence is isolated using RACE technology (Marathon Kit, ClonTech.). This approach relies on reverse transcribing mRNA into double strand cDNA, ligating linkers onto the ends of the cDNA and amplifying the desired extremity of the cDNA using a gene-specific primer and one of the linker oligonucleotides. Marathon PCR products are cloned into a plasmid (pCRII-TOPO, InVitrogen) and sequenced.

The obtained sequence (SEQ ID NO:1) has twoputative open reading frames of 193 and 262 amino acids (SEQ ID NO:2 and SEQ ID NO:3).

Example 6.

EST profiles

A complementary approach to experimental antigen tissue expression characterization is to explore the human "Expressed Sequence Tags" (ESTs) database. ESTs are small fragments of cDNA made from a collection of mRNA extracted from a particular tissue or cell line. Such database currently provides a massive amount of human ESTs (106) from several hundreds of cDNA tissue libraries, including tumoral tissues from various types and states of disease. By means of informatics tools (Blast), a comparison search of the CASB7439 sequence is performed in order to have further insight into tissue expression.

EST distribution of CASB7439 :

EST GenBank Accession number	EST cDNA tissue library
C00634	Human adult (K.Okubo)
AA468668	NCI_CGAP_Co3
AA565752	NCI_CGAP_Co11
AA565766	NCI_CGAP_Co11
AA565767	NCI_CGAP_Co11
AI337239	NCI_CGAP_Co16
AI337448	NCI_CGAP_Co16
AI393930	NCI_CGAP_CLL1
AI473673	NCI_CGAP_Co14
AI632444	NCI_CGAP_GC6
AI861937	NCI_CGAP_Co16
AI825214	NCI_CGAP_GC6
AW080652	NCI_CGAP_Co19
AW083899	NCI_CGAP_Co19
AW206058	NCI_CGAP_Sub3
AW237006	NCI_CGAP_GC6
AW364626	DT0036

AW449612	NCI_CGAP_Sub5

CASB7439 contains **9 ESTs** from **4 tumor colon** libraries, **one EST** from **one normal colon** library, **3 ESTs** from **one tumor germ cell** library, **one EST** from **one chronic lymphocyte leukemia cells** library, **2 ESTs** from **2 mixed tumors libraries**, **2 ESTs** from libraries of **unknown type**.

This clearly suggests, as expected, CASB7439 is over expressed in tumoral tissues, with an emphasis in colorectal tumoral tissues, compared to normal and colorectal normal tissues.

Example 7 :

7.1 Expression and purification of tumour-specific antigens

Expression in microbial hosts, or alternatively in vitro transcription/translation, is used to produce the antigen of the invention for vaccine purposes and to produce protein fragments or whole protein for rapid purification and generation of antibodies needed for characterization of the naturally expressed protein by immunohistochemistry or for follow-up of purification.

Recombinant proteins may be expressed in two microbial hosts, E. coli and in yeast (such as Saccharomyces cerevisiae or Pichia pastoris). This allows the selection of the expression system with the best features for this particular antigen production. In general, the recombinant antigen will be expressed in E. coli and the reagent protein expressed in yeast.

The expression strategy first involves the design of the primary structure of the recombinant antigen. In general an expression fusion partner (EFP) is placed at the N terminal extremity to improve levels of expression that could also include a region useful for modulating the immunogenic properties of the antigen, an immune fusion partner (IFP). In addition, an affinity fusion partner (AFP) useful for facilitating further purification is included at the C-terminal end.

When the recombinant strains are available, the recombinant product is characterized by the evaluation of the level of expression and the prediction of further solubility of the protein by analysis of the behavior in the crude extract.

After growth on appropriate culture medium and induction of the recombinant protein expression, total extracts are analyzed by SDS-PAGE. The recombinant proteins are visualized in stained gels and identified by Western blot analysis using specific antibodies.

A comparative evaluation of the different versions of the expressed antigen will allow the selection of the most promising candidate that is to be used for further purification and immunological evaluation.

As mentioned above, several constructs might undergo comparative evaluation :

For rapid expression and purification as well as generation of antibodies against CASB7439, it is proposed to generate in E. Coli a full length CASB7439 protein with NS1 as EFP and a histidine tail as AFP.

Therefore, two constructs are proposed :

- Full length wild type CASB7439 cDNA in fusion with NS1 cDNA as EFP and with a histidine tail coding cDNA as an AFP (SEQ ID NO:8). Fusion protein sequence is depicted in SEQ ID NO:10.
- Full length mutated CASB7439 cDNA in fusion with NS1 cDNA as EFP and with a histidine tail coding cDNA as an AFP (SEQ ID NO:9). It is proposed in this construct to have the first 50 codons of native CASB7439 cDNA replaced by codons specific of the E. Coli codon usage, to enhance expression potential of CASB7439 in its E. Coli host. Fusion protein sequence is depicted in SEQ ID NO:11.

The purification scheme follows a classical approach based on the presence of an His affinity tail in the recombinant protein. In a typical experiment the disrupted cells are filtered and the acellular extracts loaded onto an Ion Metal Affinity Chromatography (IMAC; Ni++NTA from Qiagen) that will specifically retain the recombinant protein. The retained proteins are eluted by 0-500 mM Imidazole gradient (possibly in presence of a detergent) in a phosphate buffer.

7.2 Antibody production and immunohistochemistry

Small amounts of relatively purified protein can be used to generate immunological tools in order to

a) detect the expression by immunohistochemistry in normal or cancer tissue sections;

b) detect the expression, and to follow the protein during the purification process (ELISA/ Western Blot); or

c) characterise/ quantify the purified protein (ELISA).

7.2.1 Polyclonal antibodies:

Immunization

2- 3 Rabbits are immunised , intramuscularly (I.M.) , 3 times at 3 weeks intervals with 100µg of protein, formulated in the adjuvant 3D-MPL/QS21. Three weeks after each immunisation a blood sample is taken and the antibody titer estimated in the serum by ELISA using the protein as coating antigen following a standard protocol.

ELISA

96 well microplates (maxisorb Nunc) are coated with 5µg of protein overnight at 4°C. After 1hour saturation at 37°C with PBS NCS 1%, serial dilution of the rabbit sera is added for 1H 30 at 37°C (starting at 1/10). After 3 washings in PBS Tween, anti rabbit biotinylated anti serum (Amersham) is added (1/5000). Plates are washed and peroxydase coupled streptavidin (1/5000) is added for 30 min at 37°C. After washing, 50µl TMB (BioRad) is added for 7 min and the reaction then stopped with H2SO4 0.2M. The OD can be measured at 450 nm and midpoint dilutions calculated by SoftmaxPro.

7.2.2 Monoclonal antibodies:

Immunization

5 BALB/c mice are immunized 3 times at 3 week intervals with 5 µg of purified protein. Bleedings are performed 14 days post II and 1 week post 3. The sera are tested by Elisa on purified protein used as coated antigen. Based on these results (midpoint dilution > 10000) one mouse is selected for fusion

Fusion/ HATselection

Spleen cells are fused with the SP2/0 myeloma according to a standard protocol using PEG 40% and DMSO 5%. Cells are then seeded in 96 well plates 2.5 x104 - 105 cells/well and resistant clones will be selected in HAT medium. The supernatant of these hybridomas will be tested for their content in specific antibodies and when positive, will be submitted to 2 cycles of limited dilution . After 2 rounds of screening, 3 hybridomas will be chosen for ascitis production.

7.2.3 Immunohistochemistry

When antibodies are available, immuno staining is performed on normal or cancer tissue sections, in order to determine :

- ◊ the level of expression of the antigen of the invention in cancer relative to normal tissue or
- ◊ the proportion of cancer of a certain type expressing the antigen
- ◊ if other cancer types also express the antigen
- ◊ the proportion of cells expressing the antigen in a cancer tissue

Tissue sample preparation

After dissection, the tissue sample is mounted on a cork disk in OCT compound and rapidly frozen in isopentane previously super cooled in liquid nitrogen (-160°C). The block will then be conserved at -70°C until use. 7-10µm sections will be realised in a cryostat chamber (-20, -30°C).

Staining

Tissue sections are dried for 5 min at room Temperature (RT), fixed in acetone for 10min at RT, dried again, and saturated with PBS 0.5% BSA 5% serum. After 30 min at RT either a direct or indirect staining is performed using antigen specific antibodies. A direct staining leads to a better specificity but a less intense staining whilst an indirect staining leads to a more intense but less specific staining.

7.3 Analysis of human cellular immune responses to the antigen of the invention

The immunological relevance of the antigen of the invention can be assessed by in vitro priming of human T cells. All T cell lymphocyte lines and dendritic cells are derived from PBMCs (peripheral blood mononuclear cells) of healthy donors (preferred HLA-A2 subtype). An HLA-A2.1/Kb transgenic mice is also used for screening of HLA-A2.1 peptides.

Newly discovered antigen-specific CD8+ T cell lines are raised and maintained by weekly in vitro stimulation. The lytic activity and the γ-IFN production of the CD8 lines in response to the antigen or antigen derived-peptides is tested using standard procedures.

Two strategies to raise the CD8+ T cell lines are used: a peptide-based approach and a whole gene-based approach. Both approaches require the full-length cDNA of the newly discovered antigen in the correct reading frame to be either cloned in an appropriate delivery system or to be used to predict the sequence of HLA binding peptides.

Peptide-based approach

Briefly, transgenic mice are immunized with adjuvanted HLA-A2 peptides, those unable to induce a CD8 response (as defined by an efficient lysis of peptide-pulsed autologous spleen cells) will be further analyzed in the human system.
Human dendritic cells (cultured according to Romani et al.) will be pulsed with peptides and used to stimulate CD8-sorted T cells (by Facs). After several weekly stimulations, the CD8 lines will be first tested on peptide-pulsed autologous BLCL (EBV-B transformed cell lines). To verify the proper in vivo processing of the peptide, the CD8 lines will be tested on cDNA-transfected tumour cells (HLA-A2 transfected LnCaP, Skov3 or CAMA tumour cells).

Whole gene-based approach

CD8+ T cell lines will be primed and stimulated with either gene-gun transfected dendritic cells, retrovirally transduced B7.1-transfected fibroblasts, recombinant pox virus (Kim et al.) or adenovirus (Butterfield et al.) infected dendritic cells. Virus infected cells are very efficient to present antigenic peptides since the antigen is expressed at high level but can only be used once to avoid the over-growth of viral T cells lines.

After alternated stimulations, the CD8+ lines are tested on cDNA-transfected tumour cells as indicated above. Peptide specificity and identity is determined to confirm the immunological validation.

Predicted epitopes (nonamers and decamers) binding HLA alleles :

The HLA Class I binding peptide sequences are predicted either by the Parker's algorithm (Parker, K. C., M. A. Bednarek, and J. E. Coligan. 1994. Scheme for ranking potential HLA-A2 binding peptides based on independent binding of individual peptide side-chains. J. Immunol. 152:163 and http://bimas.dcrt.nih.gov/molbio/hla_bind/) or the Rammensee method (Rammensee, Friede, Stevanovic, MHC ligands and peptide motifs: 1st listing, Immunogenetics 41, 178-228, 1995 ; Rammensee, Bachmann, Stevanovic: MHC ligands and peptide motifs. Landes Bioscience 1997, and http://134.2.96.221/scripts/hlaserver.dll/home.htm). Peptides are then screened in the HLA-A2.1/Kb transgenic mice model (Vitiello et al.).

The HLA Class II binding peptide sequences are predicted using the Tepitope algorithm, with a score cut-off set to 4 (Sturniolo, Hammer et al., Nature Biotechnology. 1999. 17;555-561).

The following tables gather the Class I and II predicted epitope sequences :

HLA-A 0201 : decamers				
Rank	**Start Position**	**Subsequence Residue Listing**	**Parker's Score °**	**SEQ ID :**
1	64	KLVNLGFQAL	142.060	SEQ ID NO:12

° : Estimate of Half Time of Disassociation of a Molecule Containing This Subsequence.

HLA-A 0201 : nonamers				
Rank	Start Position	Subsequence Residue Listing	Parker's Score °	SEQ ID :
1	182	ELLDFSSWL	507.976	SEQ ID NO:13
2	104	RLLAEHDAV	126.098	SEQ ID NO:14
3	64	KLVNLGFQA	100.850	SEQ ID NO:15

° : Estimate of Half Time of Disassociation of a Molecule Containing This Subsequence.

HLA-A 24 : nonamers				
Rank	Start Position	Subsequence Residue Listing	Parker's Score	SEQ ID :
1	97	EYIRALQRL	360.000	SEQ ID NO:16

° : Estimate of Half Time of Disassociation of a Molecule Containing This Subsequence.

HLA-A 24 : decamers				
Rank	Start Position	Subsequence Residue Listing	Parker's Score	SEQ ID :
1	97	EYIRALQRLL	360.000	SEQ ID NO:17

° : Estimate of Half Time of Disassociation of a Molecule Containing This Subsequence.

HLA-B 7 : decamers				
Rank	Start Position	Subsequence Residue Listing	Parker's Score	SEQ ID :
1	111	AVRNALAGGL	600.000	SEQ ID NO:18

° : Estimate of Half Time of Disassociation of a Molecule Containing This Subsequence.

HLA-B 4403 : decamers				
Rank	Start Position	Subsequence Residue Listing	Parker's Score	SEQ ID :
1	156	SEPGSPRSAY	360.000	SEQ ID NO:19
2	89	VETLRSAVEY	180.000	SEQ ID NO:20

° : Estimate of Half Time of Disassociation of a Molecule Containing This Subsequence.

HLA-DRB1*1501 : nonamers				
Rank	Start Position	Subsequence Residue Listing	Tepitope Score	SEQ ID :
1	99	IRALQRLLA	5.6	SEQ ID NO:21

HLA-DRB1*1502 : nonamers				
Rank	Start Position	Subsequence Residue Listing	Tepitope Score	SEQ ID :
1	99	IRALQRLLA	4.6	SEQ ID NO:21

HLA-DRB1*0402 : nonamers				
Rank	Start Position	Subsequence Residue Listing	Tepitope Score	SEQ ID :
1	120	LRPQAVRPS	5.4	SEQ ID NO:22

HLA-DRB1*1101 : nonamers				
Rank	Start Position	Subsequence Residue Listing	Tepitope Score	SEQ ID :
1	99	IRALQRLLA	4.8	SEQ ID NO:21

HLA-DRB1*1102 : nonamers				
Rank	Start Position	Subsequence Residue Listing	Tepitope Score	SEQ ID :
1	120	LRPQAVRPS	6.2	SEQ ID NO:22

HLA-DRB1*1104 : nonamers				
Rank	Start Position	Subsequence Residue Listing	Tepitope Score	SEQ ID :
1	99	IRALQRLLA	5.8	SEQ ID NO:21

HLA-DRB1*1106 : nonamers				
Rank	Start Position	Subsequence Residue Listing	Tepitope Score	SEQ ID :

1	99	IRALQRLLA	5.8	SEQ ID NO:21

HLA-DRB1*1301 : nonamers				
Rank	Start Position	Subsequence Residue Listing	Tepitope Score	SEQ ID :
1	120	LRPQAVRPS	6.6	SEQ ID NO:22
2	73	LRQHVPHGG	4.9	SEQ ID NO:23
3	31	LLRCSRRRR	4.4	SEQ ID NO:24

HLA-DRB1*1302 : nonamers				
Rank	Start Position	Subsequence Residue Listing	Tepitope Score	SEQ ID :
1	120	LRPQAVRPS	5.6	SEQ ID NO:22

HLA-DRB1*1304 : nonamers				
Rank	Start Position	Subsequence Residue Listing	Tepitope Score	SEQ ID :
1	120	LRPQAVRPS	6.2	SEQ ID NO:22
2	73	LRQHVPHGG	4.8	SEQ ID NO:23
3	31	LGFQALRQH	4.6	SEQ ID NO:24

HLA-DRB1*1305 : nonamers				
Rank	Start Position	Subsequence Residue Listing	Tepitope Score	SEQ ID :
1	99	IRALQRLLA	4.8	SEQ ID NO:21

HLA-DRB1*0703 : nonamers				
Rank	Start Position	Subsequence Residue Listing	Tepitope Score	SEQ ID :
1	112	VRNALAGGL	5.1	SEQ ID NO:25
2	98	YIRALQRLL	4.8	SEQ ID NO:26
3	65	LVNLGFQAL	4.5	SEQ ID NO:27

HLA-DRB5*0101 : nonamers				
Rank	Start Position	Subsequence Residue Listing	Tepitope Score	SEQ ID :

1	96	VEYIRALOR	4.3	SEQ ID NO:28

References

Vitiello et al. (L. Sherman), J. Exp. Med., J. Exp. Med, 1991, 173:1007-1015.

Romani et al., J. Exp. Med., 1994, 180:83-93.

Kim et al., J. Immunother., 1997, 20:276-286.

Butterfield et al., J. Immunol., 1998, 161:5607-5613.

All publications, including but not limited to patents and patent applications, cited in this specification are herein incorporated by reference as if each individual publication were specifically and individually indicated to be incorporated by reference herein as though fully set forth.

SEQUENCE INFORMATION

SEQ ID NO:1

GTACCTTGCTTTGGGGGCGCACTAAGTACCTGCCGGGAGCAGGGGGCGCACCGGGAACTCGCAGATTTCGC
CAGTTGGGCGCACTGGGGATCTGTGGACTGCGTCCGGGGGATGGGCTAGGGGGACATGCGCACGCTTTGGG
CCTTACAGAATGTGATCGCGCGAGGGGGAGGGCGAAGCGTGGCGGGAGGGCGAGGCGAAGGAAGGAGGGCG
TGAGAAAGGCGACGGCGGCGGCGCGGAGGAGGGTTATCTATACATTTAAAAACCAGCCGCCTGCGCCGCGC
CTGCGGAGACCTGGGAGAGTCCGGCCGCACGCGCGGGACACGAGCGTCCCACGCTCCCTGGCGCGTACGGC
CTGCCACCACTAGGCCTCCTATCCCCGGGCTCCAGACGACCTAGGACGCGTGCCCTGGGGAGTTGCCTGGC
GGCGCCGTGCCAGAAGCCCCCTTGGGGCGCCACAGTTTTCCCCGTCGCCTCCGGTTCCTCTGCCTGCACCT
TCCTGCGGCGCGCCGGGACCTGGAGCGGGCGGGTGGATGCAGGCGCGatggacggcggcacactgcccagg
tccgcgcccccctgcgcccccccgtcccctgtcggctgcgctgcccggcggagacccgcgtcccccggaactgtt
gcgctgcagccggcggcggcgaccggccaccgcagagaccggaggcggcgcagcggccgtagcgcggcgca
atgagcgcgagcgcaaccgcgtgaagctggtgaacttgggcttccaggcgctgcggcagcacgtgccgcac
ggcggcgccagcaagaagctgagcaaggtggagacgctgcgctcagccgtggagtacatccgcgcgctgca
gcgcctgctggccgagcacgacgccgtgcgcaacgcgctggcgggaqqqctgaggccgcaggccgtgcggc
cgtctgcgcccgcgggccgccagggaccaccccggtcgccgcctcgcccttccgcgcttcttcgtccccg
ggccgcgggggcagctcggagcccggctccccgcgttccgcctactcgtcggacgacagcggctgcgaagg
cgcgctgagtcctgcggagcgcgagctactcgacttctccagccggttaggggctactgaGCGCCCTCGA
CCTATGAGCCTCAGCCCCGGAAGCCGAGCGAGCGGCCGGCGCGCTCATCGCCGGGGAGCCCGCCAGGTGGA
CCGGCCCGCGCTCCGCCCCCAGCGAGCCGGGGACCCACCCACCACCCCCCCGCACCGCCGACGCCGCCTCGT
TCGTCCGGCCCAGCCTGACCAATGCCGCGGTGGAAACGGGCTTGGAGCTGGCCCCATAAGGGCTGGCGGCT
TCCTCCGACGCCGCCCCTCCCCACAGCTTCTCGACTGCAGTGGGGCGGGGGGCACCAACACTTGGAGATTT
TTCCGGAGGGGAGAGGATTTTCTAAGGGCACAGAGAATCCATTTTCTACACATTAACTTGAGCTGCTGGAG
GGACACTGCTGGCAAACGGAGACCTATTTTTGTACAAAGAACCCTTGACCTGGGGCGTAATAAAGATGACC
TGGACCCCTGCCCCCACTATCTGGAGTTTTCCATGCTGGCCAAGATCTGGACACGAGCAGTCCCTGACGGG
CGGGGTCCCTGGCGTGAGGCCCCCGTGACAGCCCACCCTGGGGTGGGTTTGTGGGCACTGCTGCTCTGCTA
GGGAGAAGCCTGTGTGGGGCACACCTCTTCAAGGGAGCGTGAACTTTATAAATAAATCAGTTCTGTTTAAA
AAAAAAAAAAAAAAA

SEQ ID NO:2

MDGGTLPRSAPPAPPVPVGCAARRRPASPELLRCSRRRRPATAETGGGAAAVARRNERERNRVKLVNLGFQ
ALRQHVPHGGASKKLSKVETLRSAVEYIRALQRLLAEHDAVRNALAGGLRPQAVRPSAPRGPPGTTPVAAS
PSRASSSPGRGGSSEPGSPRSAYSSDDSGCEGALSPAERELLDFSSWLGGY

SEQ ID NO:3

BC45354

MSAPAARSASGAEAHRSRALSSPLTSWRSRVARAPQDSARLRSRCRPTSRRNAGSRAPSCPRGPGTKKRGR
ARRRPGWSLAARGAQTAARPAASALPPARCARRRARPAGAAARGCTPRLSAASPPCSASCWRRRAARAAAA
PGSPSSPASRGCARAHCAALRPLRRLRSLRWPVAAAGCSATVPGTRVSAGQRSRQGRGAQGARTWAVCRRP
SRLHPPARSRSRRAAGRCRQRNRRRRGKLWRPKGASGCAPPGNSPGHAS

SEQ ID NO:4

GTACCTTGCTTTGGGGGCGCACTAAGTACCTGCCGGGAGCAGGGGGCGCACCGGGAACTCGCAGATTTCGC
CAGTTGGGCGCACTGGGGATCTGTGGACTGCGTCCGGGGGATGGGCTAGGGGGACATGCGCACGCTTTGGG
CCTTACAGAATGTGATCGCGCCGAGGGGGAGGGCCCGAAGCGTGGCGGGAGGGCGAGGCGAAGGAAGGAGGG
CGTGAGAAAGGCGACGGCGGCGGCGCGGAGGAGGGTTATCTATACATTTAAAAACCAGCCGCCTGCGCCGC
GCCTGCGGAGACCTGCCAGAGTCCGGCCGCACGCGGGGGACACGAGCGTCCCACGCTCCCTGGCGCGTACG
GCCTGCCACCACTAGGCCTCCTATCCCCGGGCTCCAGACGACCTAGGACGCGTGCCCTGGGGAGTTGCCTG
GCGGCGCCGTGCCAGAAGCCCCCTTGGGGCGCCACAGTTTTCCCCGTCGCCTCCGGTTCCTCTGCCTGCAC
CTTCCTGCGGCGCGCCGGGACCTGGAGCGGGCGGGTGGATGCAGGCGCGatggacggcggcacactgcccca
ggtccgcgcccccctgcgcccccccgtcccctgtcggctgcgctgcccggcggagacccgcgtccccggaactg
ttgcgctgcagccggcggcggcgaccggccaccgcagagaccggaggcggcgcagcggccgtagcgcggcg
caatgagcgcgagcgcaaccgcgtgaagctggtgaacttgggcttccaggcgctgcggcagcacgtgccgc
acggcggcgccagcaagaagctgagcaaggtggagacgctgcgctcagccgtgcagtacatccgcgcgctg
cagcgcctgctggccgagcacgacgccgtgcgcaacgcgctggcgggaggygctgaggccgcaggccgtgcg
gccgtctgcgccccgcgggccgccagggaccaccccggtcgccgcctcgccctcccgcgcttcttcgtccc
cgggccgcgggggcagctcggagcccggctccccgcgttccgcctactcgtcggacgacagcggctgcgaa
ggcgcgctgagtcctgcggagcgcgagctactcgacttctccagctggttaggggctactgaGCGCCCTC
GACCTAATAAGCCTCAAGCCCCGGAAACCCGAGCGAACGGGCCGGCGCGCTTCATCGCCGGGGAAGCCCGC
CAAGGTGGACCGGGCCCGCGCTCCGCCCCCAGCGAGCCGGGGACCCACCCACCACCCCCCGCACCGCCGAC
GCCGCCTCGTTCGTCCGGCCCAGCCTGACCAATGCCGCGGTGGAAACGGGCTTGGAGCTGGCCCCATAAGG
GCTGGCGGCTTCCTCCGACGCCGCCCCTCCCCACAGCTTCTCGACTGCAGTGGGGCGGGGGGCACCAACAC
TTGGAGATTTTTCCGGAGGGGAGAGGATTTTCTAAGGGCACAGAGAATCCATTTTCTACACATTAACTTGA
GCTGCTGGAGGGACACTGCTGGCAAACGGAGACCTATTTTTGTACAAAGAACCCTTGACCTGGGGCGTAAT
AAAGATGACCTGGACCCCTGCCCCCACTATCTGGAGTTTTCCATGCTGGCCAAGATCTGGACACGAGCAGT
CCCTGAGGGGCGGGGTCCCTGGCGTGAGGCCCCCGTGACAGCCCACCCTGGGGTGGGTTTGTGGGCACTGC
TGCTCTGCTAGGGAGAAGCCTGTGTGGGGCACACCTCTTCAAGGGAGCGTGAACTTTATAAATAAATCAGT
TCTGTTTAAAAAAAAAAAAAAAAAAAAAAACCGAGGGGGGGCCCGGAGCCAACAAA

SEQ ID NO:5

BC45354

GGTAAACAGAACTGATTTATTTATAAAGTTCACGCTCCCTTGAAGAGGTGTGCCCCACACAGGCTTCTCCC
TAGCAGAGCAGCAGTGCCCACAAACCCACCCCAGGGTGGGCTGTCACGGGGGCCTCACGCCAGGGACCCCG
CCCCTCAGGGACTGCTCGTGTCCAGATCTTGGCCAGCATGGAAAACTCCAGATAGTGGCCGCAGGGGTCCA
GGTCATCTTTATTACGCCCCAGGTCAAGGGTTCTTTGTACAAAAATAGGTCTCCGTTTGCCAGCAGTGTCC
CTCCAGCAGCTCAAGTTAATGTGTAGAAAATGGATTCTCTGTGCCCTTAGAAAATCCTCTCCCCTCCGGAA
AAATCTCCAAGTGTTGGTGCCCCCCGCCCCACTGCAGTCGAGAAGCTGTGGGGAGGGGCGGCGTCGGAGGA
AGCCGCAGCCCATTATGGGGCCAGCTCCAAGCCCGTTTCCACCGCGGCATTGGTCAGGCTGGGCGGACGAA
CGAGGCGGCGTCGCCCGTGCCGCGGCGTCGTCGCTCGGTCCCCGGCTCGCTGGGGGCGGAGCAGCGGGCCGG
TCCACCTGGCGGGCTCCCC

SEQ ID NO:6

MDGGTLPRSAPPAPFVPVGCAARRRPASPELLRCSRRRRPATAETGGGAAAVARRNERERNRVKLVNLGFQ
ALRQHVPHGGASKKLSKVETLRSAVEYIRALQRLLAEHDAVRNALAGGLRPQAVRPSAPRGPPGTTPVAAS
PSRASSSPGRGGSSEPGSPRSAYSSDDSGCEGALSPAERELLDFSSWLGGY

SEQ ID NO:7

MCRKWLLCALRKSSPLRKNLQVLVPPAPLQSRSCGEGRRRRKPPALMGPAPSPFPPRHWSGWAGRTRRRRR
GGGWWVGPRLAGGGARARSTLAGFPGDEARRPVRSGFRGLRLIRSRALSSPLTSWRSRVARAPQDSARLRS
RCRPTSRRNAGSRAPSCPRGPGTKKRGRARRRPGWSLAARGAQTAARPAASALPPARCARRRARPAGAAAR
GCTPRLSAASPPCSASCWRRRAARAAAAPGSPSSPASRGCARAHCAALRPLRRLRSLRWPVAAAGCSATVP
GTRVSAGQRSRQGRGAQGARTWAVCRRPSRLHPPARSRSRRAAGRCRQRNRRRRGKLWRPKGASGTAPPGN
SPGHAS

SEQ ID NO:8

ATGGATCCAAACACTGTGTCAAGCTTTCAGGTAGATTGCTTTCTTTGGCATGTCCGCAAACGAGTTGCAGA
CCAAGAACTAGGTGATGCCCCATTCCTTGATCGGCTTCGCCGAGATCAGAAATCCCTAAGAGGAAGGGGCA
GCACCCTCGGTCTGGACATCGAGACAGCCACACGTGCTGGAAAGCAGATAGTGGAGCGGATTCTGAAAGAA
GAATCCGATGAGGCACTTAAAATGACCATGGACGGCGGCACACTGCCCAGGTCCGCGCCCCCTGCGCCCCC
CGTCCCTGTCGGCTGCGCTGCCCGGCGGAGACCCGCGTCCCCGGAACTGTTGCGCTGCAGCCGGCGGCGGC
GACCGGCCACCGCAGAGACCGGAGGCGGCGCAGCGGCCGTAGCGCGGCGCAATGAGCGCGAGCGCAACCGC
GTGAAGCTGGTGAACTTGGGCTTCCAGGCGCTGCGGCAGCACGTGCCGCACGGCGGCGCCAGCAAGAAGCT
GAGCAAGGTGGAGACGCTGCGCTCAGCCGTGGAGTACATCCGCGCGCTGCAGCGCCTGCTGGCCGAGCACG
ACGCCGTGCGCAACGCGCTGGCGGGAGGGCTGAGGCCGCAGGCCGTGCGGCCGTCTGCGCCCCGCGGGCCG
CCAGGGACCACCCCGGTCGCCGCCTCGCCCTCCCGCGCTTCTTCGTCCCCGGGCCGCGGGGGCAGCTCGGA

GCCCGGCTCCCCGCGTTCCGCCTACTCGTCGGACGACAGCGGCTGCGAAGGCGCGCTGAGTCCTGCGGAGC
GCGAGCTACTCGACTTCTCCAGCTGGTTAGGGGGCTACACTAGTGGCCACCATCACCATCACCATTAA

SEQ ID NO:9

ATGGATCCAAACACTGTGTCAAGCTTTCAGGTAGATTGCTTTCTTTGGCATCTCCGCAAACGAGTTGCAGA
CCAAGAACTAGGTGATGCCCCATTCCTTGATCGGCTTCGCCGAGATCAGAAATCCCTAAGAGGAAGGGGCA
GCACCCTCGGTCTGGACATCGAGACAGCCACACGTGCTGGAAAGCAGATAGTGGAGCGGATTCTGAAAGAA
GAATCCGATGAGGCACTTAAAATGACCATGGACGGCGGCACCCTGCCGCGTTCCGCGCCGCCGGCGCCGCC
AGTTCCGGTTGGCTGCGCTGCCCGTCGCCGTCCCGCGTCCCCGGAACTGCTGCGCTGCAGCCGTCGCCGTC
GCCCGGCCACCGCAGAGACCGGAGGCGGCGCAGCGGCCGTAGCGCGGCGCAATGAGCGCGAGCGCAACCGC
CTGAAGCTGGTGAACTTGGGCTTCCAGGCGCTGCGGCAGCACGTGCCGCACGGCGGCGCCAGCAAGAAGCT
GAGCAAGGTGGAGACGCTGCGCTCAGCCGTGGAGTACATCCGCGCGCTGCAGCGCCTGCTGGCCGAGCACG
ACGCCGTGCGCAACGCGCTGGCGGGAGGGCTGAGGCCGCAGGCCGTGCGGCCGTCTGCGCCCCGCGGGCCG
CCAGGGACCACCCCGGTCGCCGCCTCGCCCTCCCGCGCTTCTTCGTCCCCGGGCCGCGGGGGCAGCTCGGA
GCCCGGCTCCCCGCGTTCCGCCTACTCGTCGGACGACAGCGGCTGCGAAGGCGCGCTGAGTCCTGCGGAGC
GCGAGCTACTCCACTTCTCCAGCTGGTTAGGGGGCTACACTAGTGGCCACCATCACCATCACCATTAA

SEQ ID NO:10

MDPNTVSSFQVDCFLWHVRKRVADQELGDAPFLDRLRRDQKSLRGRGSTLGLDIETATRAGKQIVERILKE
ESDEALKMTMDGGTLPRSAPPAPPVPVGCAARRRPASPELLRCSRRRRPATAETGGGAAAVARRNERERNR
VKLVNLGFQALRQHVPHGGASKKLSKVETLRSAVEYIRALQRLLAEHDAVRNALAGGLRPQAVRPSAPRGP
PGTTPVAASPSRASSSPGRGGSSEPGSPRSAYSSDDSGCEGALSPAERELLDFSSWLGGYTSGHHHHHH

SEQ ID NO:11

MDPNTVSSFQVDCFLWHVRKRVADQELGDAPFLDRLRRDQKSLRGRGSTLGLDIETATRAGKQIVERILKE
ESDEALKMTMDGGTLPRSAPPAPPVPVGCAARRRPASPELLRCSRRRRPATAETGGGAAAVARRNERERNR
VKLVNLGFQALRQHVPHGGASKKLSKVETLRSAVEYIRALQRLLAEHDAVRNALAGGLRPQAVRPSAPRGP
PGTTPVAASPSRASSSPGRGGSSEPGSPRSAYSSDDSGCEGALSPAERELLDFSSWLGGYTSGHHHHHH

SEQ ID NO:12

KLVNlGFQAL

SEQ ID NO:13

ELLDFGSWL

SEQ ID NO:14

RLLAEHDAV

SEQ ID NO:15

KLVNLGFQA

SEQ ID NO:16

EYIRALQRL

SEQ ID NO:17

EYIRALQRLL

SEQ ID NO:18

AVRNALAGGL

SEQ ID NO:19

SEPGSPRSAY

SEQ ID NO:20

VETLRSAVEY

SEQ ID NO:21

IRALQRLLA

SEQ ID NO:22

BC45354

LRPQAVRPS

SEQ ID NO:23

LRQHVPHGG

SEQ ID NO:24

LGFQALRQH

SEQ ID NO:25

VRNALAGGL

SEQ ID NO:26

YIRALQRLL

SEQ ID NO:27

LVNLGFQAL

SEQ ID NO:28

VEYIRALQR

Claims

1. An isolated polypeptide comprising an amino acid sequence which has at least 70% identity to the amino acid sequence of SEQ ID NO:2 or SEQ ID NO:3 over the entire length of SEQ ID NO:2 or SEQ ID NO:3 respectively.

2. An isolated polypeptide as claimed in claim 1 in which the amino acid sequence has at least 95% identity to SEQ ID NO:2 or SEQ ID NO:3.

3. The polypeptide as claimed in claim 1 comprising the amino acid sequence of SEQ ID NO:2 or SEQ ID NO:3.

4. The isolated polypeptide of SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:6, or SEQ ID NO:7.

5. A polypeptide comprising an immunogenic fragment of a polypeptide as claimed in any one of claims 1 to 4 in which the immunogenic activity of the immunogenic fragment is substantially the same as the polypeptide of SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:6, or SEQ ID NO:7.

6. A polypeptide as claimed in any of claims 1 to 5 wherein said polypeptide is part of a larger fusion protein.

7. A polypeptide as claimed in any of claims 1 to 6 chemically conjugated to a carrier protein.

8. An isolated polynucleotide encoding a polypeptide as claimed in any of claims 1 to 6.

9. An isolated polynucleotide comprising a nucleotide sequence encoding a polypeptide that has at least 70% identity to the amino acid sequence of SEQ ID NO:2 or SEQ ID NO:3, over the entire length of SEQ ID NO:2 or SEQ ID NO:3 respectively; or a nucleotide sequence complementary to said isolated polynucleotide.

10. An isolated polynucleotide comprising a nucleotide sequence that has at least 70% identity to a nucleotide sequence encoding a polypeptide of SEQ ID NO:2 or SEQ ID NO:3 respectively, over the entire coding region; or a nucleotide sequence complementary to said isolated polynucleotide.

11. An isolated polynucleotide which comprises a nucleotide sequence which has at least 70% identity to that of SEQ ID NO:1 over the entire length of SEQ ID NO:1; or a nucleotide sequence complementary to said isolated polynucleotide.

12. The isolated polynucleotide as defined in any one of claims 8 to 11 in which the identity is at ~~least~~ 95%.

13. An isolated polynucleotide selected from:

(a) a polynucleotide comprising a nucleotide sequence encoding the polypeptide of SEQ ID NO:2 or SEQ ID NO:3;

(b) the polynucleotide of SEQ ID NO:1; and

(c) a polynucleotide obtainable by screening an appropriate library under stringent hybridization conditions with a labeled probe having the sequence of SEQ ID NO:1 or a fragment thereof said polynucleotide encoding a protein which has similar immunogenic properties to those of the protein of sequence ID NO:2 or SEQ ID NO:3 or a nucleotide sequence complementary to said isolated polynucleotide

14. An expression vector or a recombinant live microorganism comprising an isolated polynucleotide according to any one of claims 8 - 13.

15. A host cell comprising the expression vector of claim 14 or the isolated polynucleotide of claims 8 to 13.

16. A process for producing a polypeptide of claims 1 to 7 comprising culturing a host cell of claim 15 under conditions sufficient for the production of said polypeptide and recovering the polypeptide from the culture medium.

17. A vaccine comprising an effective amount of the polypeptide of any one of claims 1 to 7 and a pharmaceutically acceptable carrier.

18. A vaccine comprising an effective amount of the polynucleotide of any one of claims 8 to 13 and a pharmaceutically effective carrier.

19. A vaccine comprising an effective amount of antigen presenting cells, modified by in vitro loading with a polypeptide of any one of claims 1 to 7, or genetically modified in vitro to express a polypeptide of claims 1 to 7 and a pharmaceutically effective carrier.

20. A vaccine as claimed in any one of claims 17 to 19 which additionally comprises a TH-1 inducing adjuvant.

21. A vaccine as claimed in claim 20 in which the TH-1 inducing adjuvant is selected from the group of adjuvants comprising: 3D-MPL, QS21, a mixture of QS21 and cholesterol, and a CpG oligonucleotide.

22. An antibody immunospecific for the polypeptide or immunological fragment as claimed in any one of claims 1 to 5.

23. A method for screening to identify compounds which stimulate or which inhibit the function of the polypeptide of any one of claims 1 to 5 which comprises a method selected from the group consisting of:

(a) measuring the binding of a candidate compound to the said polypeptide (or to the cells or membranes bearing the polypeptide) or a fusion protein thereof by means of a label directly or indirectly associated with the candidate compound;

(b) measuring the binding of a candidate compound to the said polypeptide (or to the cells or membranes bearing the polypeptide) or a fusion protein thereof in the presence of a labeled competitior;

(c) testing whether the candidate compound results in a signal generated by activation or inhibition of the said polypeptide, using detection systems appropriate to the cells or cell membranes bearing the polypeptide;

(d) mixing a candidate compound with a solution containing a polypeptide of any one of claims 1 to 7, to form a mixture, measuring activity of the polypeptide in the mixture, and comparing the activity of the mixture to a standard; or

(e) detecting the effect of a candidate compound on the production of mRNA encoding said polypeptide and said polypeptide in cells, using for instance, an ELISA assay.

24. A method for the treatment of a subject by immunoprophylaxis or therapy comprising *in vitro* induction of immune responses to a molecule of any one of claims 1 to 5, using *in vitro* incubation of the polypeptide of any one of claims 1 to 7 or the polynucleotide of any one of claims 8 to 13 with cells from the immune system of a mammal, and reinfusing these activated immune cells to the mammal for the treatment of disease.

25. A method as claimed in claim 24 wherein the treatment is for colon cancer.

26. An agonist or antagonist to the polypeptide of claims 1 to 5.

27. A compound which is:

(a) an agonist or antagonist to the polypeptide of claims 1 to 5;

(b) isolated polynucleotide of claims 8 to 13; or

(c) a nucleic acid molecule that modulates the expression of the nucleotide sequence encoding the polypeptide of any one of claims 1 to 5;

for use in therapy.

28. A process for diagnosing a disease or a susceptibility to a disease in a subject related to expression or activity of a polypeptide of any one of claims 1 to 5 in a subject comprising analyzing for the presence or amount of said polypeptide in a sample derived from said subject.

29. A process for diagnosing a disease or a susceptibility to a disease in a subject related to expression or activity of a polynucleotide of any one of claims 8 to 13 in a subject comprising analyzing for the presence or amount of said polynucleotide in a sample derived from said subject.

30. A process for diagnosing the presence of colon cancer or a susceptibility to colon cancer in a subject related to expression or activity of a polypeptide of any one of claims 1 to 5 in a subject comprising analyzing for the presence or amount of said polypeptide in a sample derived from said subject.

31. A process for diagnosing the presence of colon cancer or a susceptibility to colon cancer in a subject related to expression or activity of a polynucleotide of any one of claims 8 to 13 in a subject comprising analyzing for the presence or amount of said polynucleotide in a sample derived from said subject.

32. An isolated polynucleotide selected from the group consisting of:
(a) an isolated polynucleotide comprising a nucleotide sequence which has at least 70% identity to SEQ ID NO:4 or SEQ ID NO:5 over the entire length of SEQ ID NO:4 or SEQ ID NO:5 respectively;
(b) an isolated polynucleotide comprising the polynucleotide of of SEQ ID NO:4 or SEQ ID NO:5;
(c) the polynucleotide of of SEQ ID NO:4 or SEQ ID NO:5.

33. A live vaccine composition comprising an expression vector or recombinant live micro-organism according to claim 14.

34. Use of a polynucleotide as claimed in any one of claims 8 to 13 for the manufacture of a medicament in the treatment of carcinoma.

35. Use of a polynucleotide as claimed in any one of claims 8 to 13 for the manufacture of a medicament in the treatment of colon carcinoma.

36. Use of a polypeptide as claimed in any one of claims 1 to 7 for the manufacture of a medicament in the treatment of carcinoma.

37. Use of a polypeptide as claimed in any one of claims 1 to 7 for the manufacture of a medicament in the treatment of colon carcinoma.

Abstract

CASB7439 polypeptides and polynucleotides and methods for producing such polypeptides by recombinant techniques are disclosed. Also disclosed are methods for utilizing CASB7439 polypeptides and polynucleotides in diagnostics, prophylaxis and therapy, and diagnostic assays for such.

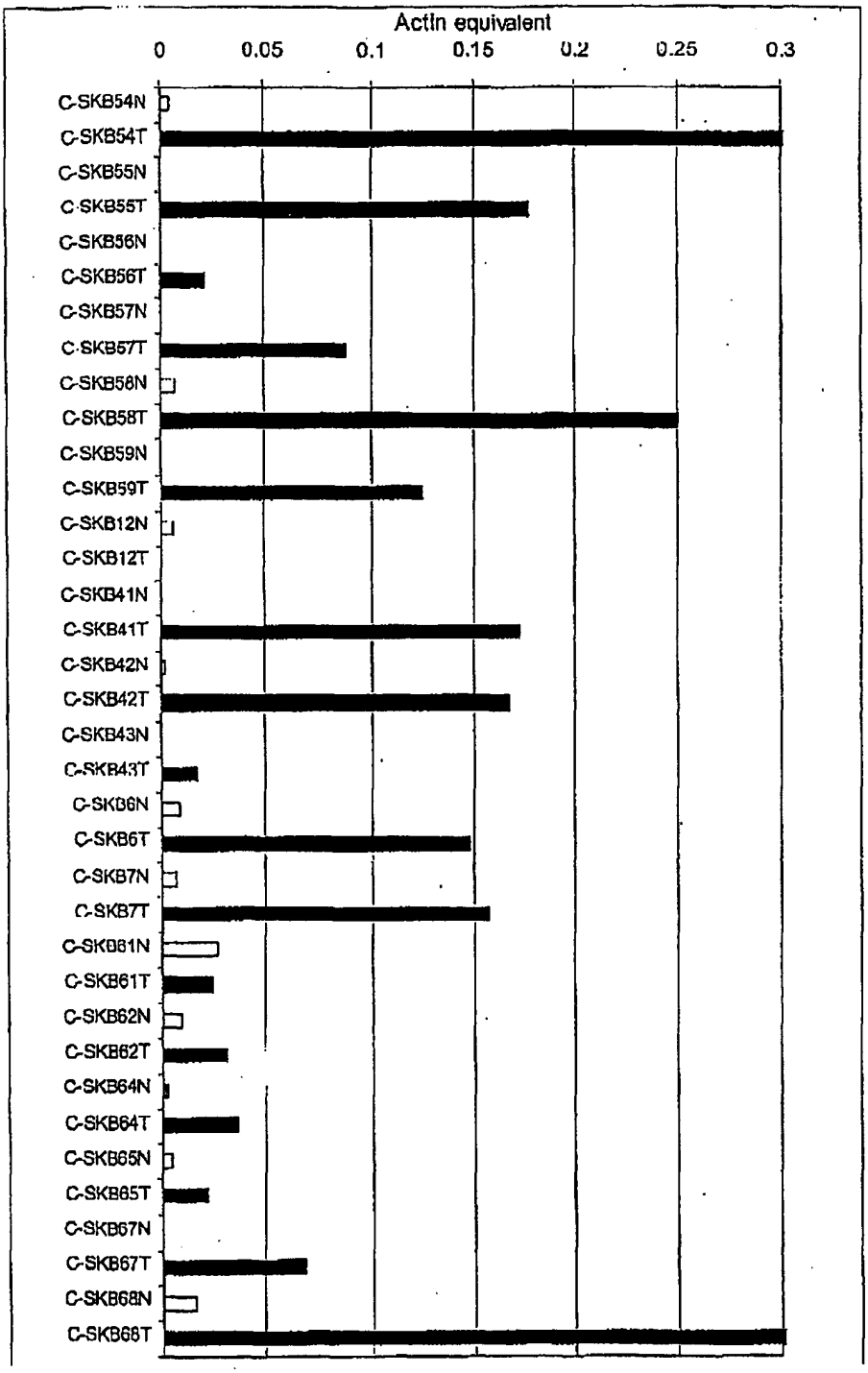


Figure 1: RT PCR data of CASB7439 transcript on coloractal tumours and matched normal colon.

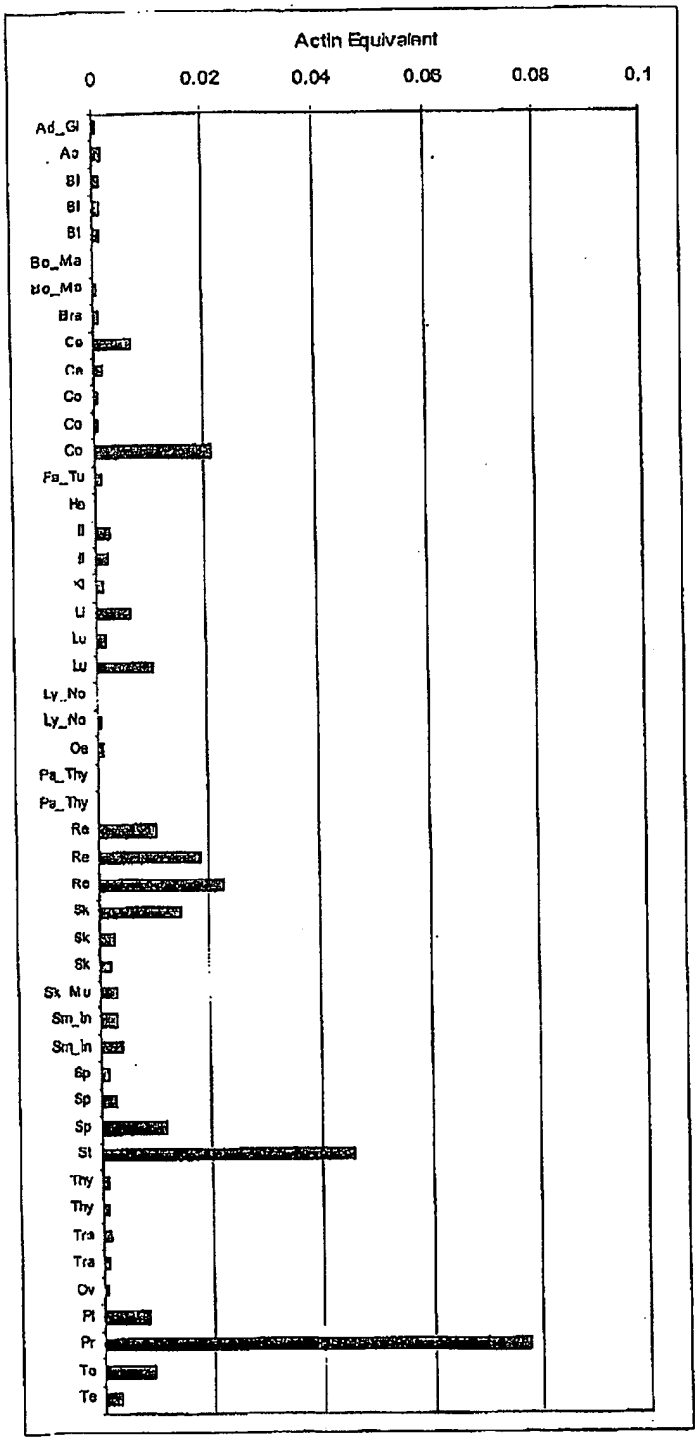


Figure 2: RT PCR data of CASB7439 transcript in a panel of normal tissues.

Legend:
Ad_Gl: adrenal gland;
Ao: aorta; Bl: bladder,
Bo_Ma: bone marrow;
Bra: brain;
Ce: cervix; Co: colon;
Fa_Tu: fallopian tube; He: heart;
Il: ileon; Ki: kidney; Li: liver;
Lu: lung; Ly_No: lymph node;
Oe: oesophagus; Ov: ovary;
Pa_Thy: parathyroid gland;
Pl: placenta; Pr: prostate;
Re: rectum; Sk: skin;
Sk_Mu: skeletal muscle;
Sm_In: small intestine ;
Sp: spleen; St: stomach; Te: testis;
Thy: thyroid gland; Tra: trachea;
Bre: Breast.